DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS

Editors: Matthew French, Simon Jackson, Elina Jokisuu

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS

interdisciplinary
graduate conference
2010

Proceedings of the University of Cambridge Interdisciplinary Graduate Conference
June 2010, Cambridge, United Kingdom

EDITED BY

MATTHEW FRENCH
Department of Architecture, University of Cambridge, UK

SIMON JACKSON
English Faculty, University of Cambridge, UK

ELINA JOKISUU
Department of Engineering, University of Cambridge, UK

+

Graduate Development Programme

Published by the Graduate Development Programme, University of Cambridge
25 Trumpington Street, Cambridge, CB2 1QA, United Kingdom

First Edition (CD):
ISBN: 978-0-9566139-0-5

Second, revised edition (Online, D-Space, University of Cambridge Library):
ISBN: 978-0-9566139-1-2

Diverse Engagement: Drawing in the Margins was organized by a committee of five Cambridge University graduate students, in alphabetical order: Matthew French, Simon Jackson, Elina Jokisuu, Rachelle Larocque and Jie Pan with support from the Graduate Development Programme.

CONTENTS

FOREWORD

Introduction

You are reading the proceedings of the Interdisciplinary Graduate Conference 2010. The Interdisciplinary Graduate Conference (IGC), now in its third year, was created by the Graduate Development Programme at the University of Cambridge to provide postgraduate researchers from all fields of study a platform for collaboration and an opportunity to exchange ideas and experiences. At the time of writing, the IGC 2010 is shaping up to be a huge success. We received over 130 submitted abstracts from which we selected 72 papers for oral presentation and 20 poster presentations. So far the conference has over 140 registered delegates and a fantastic line-up of six keynote speakers. It promises to be an exciting and engaging two-day conference.

Adding to the success of the IGC 2010, this is the first time that delegates have had the opportunity to publish a full paper in conference proceedings. The committee decided to publish proceedings to give delegates the opportunity to publish their work as well as provide a permanent record of the conference. This collection of 31 papers highlights the extraordinary diversity of research topics. Each paper reflects in some way the theme of the conference, *Diverse Engagement: Drawing in the Margins*. The theme was originally conceived to allow a wide variety of interpretations of what a margin means. It might mean working in the margins of academia or at the juncture of different disciplines. It could be interpreted as research that challenges the dominant theories and paradigms or research that involves marginalised communities. Or, the margin could simply mean a place to doodle to get the creative juices flowing. Whatever the meaning of the margin, each of the papers demonstrates the importance of interdisciplinary research: looking at existing theories through the lenses of another discipline reveals new perspectives and unanswered questions.

We hope you will find these proceedings as inspiring and engaging as we do.

The Editors

Acknowledgements

The Editors would like to acknowledge the following individuals and organizations who helped to make this year's conference possible:

First and foremost, we would like to thank Rachelle Larocque and Jie Pan who have been valuable members of the IGC 2010 organizing committee.

Our keynote speakers, Prof. David Spiegelhalter, Dr Sheila Preston, Dr Allan McRobie, Dr Ian Wilson, Dr Ben Crewe, and Prof. Julie Sanders

For financial support and sponsorship, we would like to thank the Schools of the Humanities and Social Sciences, the Arts and Humanities and the School of the Physical Sciences, University of Cambridge; and the Masters and Fellows of the following colleges: Churchill, Clare Hall, Emmanuel, Fitzwilliam, Newnham, Queens', St Catherine's, St John's, Selwyn, Sidney Sussex, and Trinity.

Jeff Vickers, for help with the website; Bernhard Dusch for graphic design; and Stephen Lu for his helpful contributions in organizing the conference.

The committee would also like to the Faculty of Engineering, University of Cambridge, for hosting the conference and Selwyn College for hosting the conference dinner.

Finally, to thank Katie Hewitt and Dr. Jeremy Schildt of the University's Graduate Development Programme, for their support and advice in running this year's conference.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

CEREMONY: A CASE STUDY IN LITERARY ANTHROPOLOGY

Francesca ANIBALLI

PhD Student, University of Glasgow, Scotland, UK
f.aniballi.1@research.gla.ac.uk

Keywords: Anthropology, Literature, Ritual, Magic, Magical Realism, Etic, Emic

ABSTRACT

This paper is an experiment in literary anthropology inasmuch as it attempts to apply an anthropological perspective to the literary category of 'magical realism' through the example of a specific case study. Anthropology of literature can provide suitable tools for going beyond the Western bias for mimesis, thus opening up new terrain for literary criticism.

Starting from an exploration of anthropological ideas of 'magic' and 'ritual', it goes on to show how *Ceremony*, a novel by Native American author Leslie Marmon Silko, published in 1977, is an example of 'performative' literature. It also contrasts some of the Native American cultural material inside the novel with E. C. Parsons' anthropological account of a specific ritual and points out the anthropological distinction between 'emic' and 'etic' perspectives, in order to analyze 'magical realism' in the text.

By relating the novel to the author's declarations on one side, and to Parsons' anthropology as 'science' on the other, the paper shows the limits of Western discourses of 'objectivity' when interpreting 'magical realism' in the text. It is in fact evident that 'magical realism' is an ambiguous label for what the author would regard as her way of representing or, rather, performing 'reality' according to mythical patterns of communal significance.

Because it situates itself on the border between Native American communal worldview and Western readers' expectations of a journey into the unknown and the unfamiliar, the novel gains greatly by a reading which incorporates the 'emic' – 'etic' tension, while finally providing a 'healing ritual' for both sides of the border. It is evident that the idea that 'reality' can be recreated by the re-enactment of the community's ancient myths and that literature participates to this performative function is at the very heart of the novel.

INTRODUCTION

In their introduction to the volume *Magical realism: theory, history, community*, editors Zamora and Faris state:

> *An essential difference […] between realism and magical realism involves the intentionality implicit in the conventions of the two modes. […] Realism intends its version of the world as a singular version, as an objective (hence universal) representation of natural and social realities – in short, that realism functions ideologically and hegemonically. Magical realism also functions ideologically but […] less hegemonically, for its program is not centralizing but eccentric: it creates space for interactions of diversity. […] Magical realism often facilitates the fusion, or coexistence, of possible worlds, spaces, systems that would be irreconcilable in other modes of fiction. […] So magical realism may be considered an extension of realism in its concern with the nature of reality and its representation, at the same time that it resists the basic assumptions of post-enlightenment rationalism and literary realism.*[1]

ISBN 978-0-9566139-1-2

Indeed all the essays contained in the volume weave a rich comparative tapestry of different 'magical realisms', so that Connell's fear that 'magical realism' can be used as an essentialist label to reduce all post-colonial literature to the state of pre-modern expression of a 'primitive' or 'traditional' culture is partly misplaced.[2]

Yet, there are some points in his critique that are worth considering. He argues that magical realist texts should be included in Modernism and not opposed to it, on the basis of the fact that both types of literature deal with modernization and negotiate the changes brought about by it. Although in this paper I am focusing neither on modernism nor postmodernism, I think that Connell's justified preoccupation about formal descriptions of magic realism as 'characterized by a juxtaposition of apparently reliable, realistic reportage and extravagant fantasy',[3] can be eased through attempting to broach a new perspective and throw new light on the old issue of definition.

For this purpose, I propose that the anthropological concepts of 'emic' and 'etic' viewpoints are essential to a correct understanding of 'magical realism', as not only an 'extension' of realism but an alternative way to write about 'reality', one which corresponds to the emic point of view of the narrative voice within a novel. It is my contention that 'magical realism' can be a sort of 'performative' literature, rather than a 'mimetic' one and am going to illustrate a specific case through the analysis of the novel *Ceremony*, published in 1977 by Native American author Leslie Marmon Silko.

In order to explain literature as 'performative', I will rely on Tambiah's anthropological ideas of 'magic' and 'ritual', which will help show my point in the novel. Before moving on, it is worthwhile to pick up Connell's remarks about ritual and the nature of magic. Following 'Turner's explanation of "rituals"', Connell sees them as an 'active way of negotiating new situations,' such as 'modernization'[4] and later he reports Lèvi-Strauss's view that 'science and magic…require the same sort of mental operations and they differ not so much in kind as in the different types of phenomena to which they are applied'.[5] Thus Connell debunks those definitions of 'magical realism' based on 'the dissimilarity of the two modes of thinking',[6] the mythical and the scientific. He also points to the ambiguous location of anthropology as a discipline that conflates the 'scientific' and 'myth'. However, far from being negative, this liminal status of anthropology is what makes it a suitable field and set of methodologies for inquiry onto literature too. Thus, I am advocating the bridging of anthropology and (comparative) literature into anthropology of literature. Let us now turn to Tambiah's ideas of 'magic' and 'ritual'.

DISCUSSION

Magic and Ritual

According to anthropologist Tambiah, magic is neither true nor false. Magic is a performance: a series of actions carefully planned and carried out in a certain way, capable of re-enacting the play of the forces which animate the universe according to a culture-specific vision. Therefore magic proper cannot be understood if it is not placed within the wider framework of communal cultural practices such as rituals. Rituals are the practical actions that correspond to the illocutionary force of magical formulas and words. In Tambiah's words:

> *[…] we can say that ritual acts and magical rites are of the "illocutionary" or "performative" sort, which simply by virtue of being enacted […] achieve a change of state, or do something effective (for example, an installation ceremony undergone by the candidate makes him a "chief"). This performative aspect of the rite should be distinguished from its locutionary (referential, information-carrying) and perlocutionary (consequences for the participants) features.*[7]

A living body of magic can be understood only within the wider scope of a society's customs and culture: together with beliefs in the supernatural, also a society's legal, political and social conventions are part of its cosmology. Thus, when dealing with ritual, '*the traditional distinction between religious and secular is of little relevance […] anything toward which an "unquestioned" and "traditionalizing" attitude is adopted can be viewed as sacred.*'[8]

Tambiah suggests that magic cannot be judged from a 'scientific' or 'objective' point of view as in it words are actions and actions are words. Therefore magic is a 'rhetorical art'[9] and belongs to the symbolic and performative realms. Moreover, most anthropologists agree in considering ritual as inherently conservative in its social functions: it re-instates the belonging of the individual to the community and the predominance of communal values over individuals' interests.

ISBN 978-0-9566139-1-2

Ceremony and magical realism

Ceremony, a novel published in 1977 by Native American author Leslie Marmon Silko, draws its narrative materials from the traditional stories of the Laguna Pueblo in New Mexico, where the author grew up. Therefore it is particularly suited to be discussed as an example of performative literature, that is, a novel which takes the readers in and, as it unfolds, makes them take part to the healing ritual that the protagonist Tayo undergoes. Tayo, like the author, is the child of a Laguna woman and an unknown white man. As such, he is already set apart as the living proof of the shame his mother has brought upon her family and her village as a whole. In spite of having always been made aware of this burden by his aunt, it is only when he comes back from World War II that he realizes he has ether to be part of the community or drift away in some hospital for veterans or end up derelict somewhere on the Highway 66, in the city of Gallup, where all the Indians who try to live in the whites' world end up as drunkards and prostitutes.

Together with his personal plight and the depiction of the harsh conditions awaiting those Indians who have gone to the schools in the reservation and have been taught by white teachers to despise their native 'superstitions', the novel also spins and interweaves traditional Laguna stories throughout, in such a way that they correspond mirror-like to the events in Tayo's life and his surroundings. Hence the personal dimension of Tayo's story is being constantly re-enacted on a mythical level, by analogy, through the various myths interwoven throughout the story. Poetry and prose alternate and point to these two levels of meaning. The readers are made privy to the mythical dimension only in part, as the 'strangeness' of the Laguna mythical characters is preserved through their native names.

It is a fact that while the Laguna medicine man, who first attempts to heal Tayo from his anguish and trauma, speaks only his native language, the second medicine ma, the far more powerful Betonie is a Navajo with Mexican blood and speeks good English. He is the one who points to the need for change in order to devise new, effective ceremonies: the old ones, in fact, are too far behind. Ceremonies too have to change in order to counteract a more powerful evil than that known in the old days.

From Tayo's encounter with Betonie to the end, the readers follow and take part in the protagonist own attempt to make things whole again for himself and his people affected by the drought. Thus once again, individual wellbeing and communal welfare are deeply entangled and interconnected.

The women Tayo meets in his life are also no ordinary ones. The first is his uncle's lover, Night Swan, a powerful, self-aware Mexican dancer with catlike eyes. She gives him an important teaching:

> *Indians or Mexicans or whites – most people are afraid of change. They think if their children have the same color of skin, the same color of eyes, that nothing is changing… They are fools. They blame us, the ones who look different. That way they don't have to think about what has happened inside themselves.*[10]

The second, T'seh, is mostly called only 'the woman', as to reinforce her mythical character and teaches Tayo about love, the magical properties of herbs and stones, how to regain his uncle's lost cattle and avoid his enemy Emo's 'witchery'. T'seh, the hunter's wife is an embodiment of the mythical fertility and summer goddess, whose husband symbolizes the north and the winter, while Tayo takes on the role of the south and summer: she spends six months with each of them, according to the myth. The novel shows the intricacies of life, its small and big interconnected patterns, the need for tradition as well as renewal and change.

Towards the end of the novel Tayo has an epiphany:

> *He cried the relief he felt at finally seeing the pattern, the way all stories fit together – the old stories, the war stories, their stories – to become the story that was still being told. He was not crazy; he had never been crazy. He had only seen and heard the world as it always was: no boundaries, only transitions through all distances and time.*[11]

Indeed the novel displays Tayo's point of view throughout and attempts to convey this very vision, a world of 'transitions'. At this point it is necessary to turn to some contributions regarding the literary concept of 'magical realism'.

In his essay 'Magic realism as post-colonial discourse', Stephen Slemon declares:

> *The term "magical realism" is an oxymoron, one that suggests a binary opposition between the representational code of realism and that, roughly, of fantasy. In the language of narration in a magic realist text, a battle between two*

ISBN 978-0-9566139-1-2

> *oppositional systems takes place, each working toward the creation of a different kind of fictional world from the other. Since the ground rules of these two worlds are incompatible, neither one can fully come into being, and each remains suspended, locked in a continuous dialectic with the "other", a situation which creates disjunction within each of the separate discursive systems, rending them with gaps, absences and silences.*[12]

Now, as we have seen, it is this very binary logic that *Ceremony* calls into question throughout its development: mythical and personal planes and Indian and white worlds are so intertwined, that in spite of the destruction involved in the culture contact, everything needs changing and updating while staying true to a cosmic pattern. Stories and ceremonies need to keep true to their functions and their relations to each other, while changing and updating. Thus the medicine man Betonie piles up stacks of telephone books, calendars and newspapers in order to '*keep track of things*'[13], while Tayo knows that his '*sickness was only part of something larger, and his cure would be found only in something great and inclusive of everything.*'[14]

Betonie says that '*the ceremonies have always been changing*,'[15] thus pointing to the fact that tradition can only exist in an ever-changing flux and in negotiating with this reality. It is interesting at this point to keep in mind Tambiah's ideas concerning magic and ritual, seen as being made up of a semantic side and a pragmatic one. This means that while the magical act is being performed as a part of the ritual there are two frames of reference to take into account. First, in ritual magical utterances are accompanied by objects that function as mediating substances for a transfer of properties from an object to another (a thing, an animal or a person) in need of them, through the process of metaphor (analogical substitution) and metonymy.[16] Second, each ritual is an action and as such is part of a continuum of actions: it takes place in a wider context of so-called ordinary activities, social or technical.[17]

In *Ceremony* this principle is at work all the time: all the while, entwined with the healing ceremony, Tayo has been involved in the search for his uncle's cattle: healing ritual and practical activity are closely linked. The readers become onlookers in this ritual-novel against illness and death brought about by 'witchery'. In fact it's only when Tayo meets the human embodiment of the goddess of fertility and overcomes the temptation to react against Emo's 'witchery' that he completes his own healing process and is also re-included in the community through a communal event in the village, where the elders question him and recognize that he has really met the goddess who brings healing, blessings and rain. Tayo, the different one, the 'half-bred' turns out to be the new lymph in an ever-returning cycle and embodies the renewal of tradition itself. In *Ceremony* there is no 'binary opposition' or 'battle' between two 'incommensurable' worlds, rather Tayo's vision at the end of the novel is one of the unity of all things and of all stories. Therefore Slemon's suggestions are not universally valid when dealing with so-called 'magical realist' texts.

The other author I would like to mention is William Spindler, who in his article 'Magic realism: a typology'[18] suggests that there are three kinds of magic realism, which however, are by no means incompatible: European 'metaphysical' magic realism, with its sense of estrangement and the uncanny, exemplified by Kafka's fiction, 'ontological' magical realism, characterized by 'matter-of-factness' in relating 'inexplicable' events, and 'anthropological' magical realism, where a Native worldview is set side by side to the Western rational worldview. This definition is still tainted by Western biases, not least a bias for mimesis, art as imitation of life and nature. So far I have tried to show how in a specific novel, *Ceremony*, Tambiah's ideas on ritual can be useful to draw the outlines of an alternative concept for describing a highly original magical realist text: that of the performative. Literature as 'performance' rather than only mimesis: mimesis is included in the performance, which, however, goes beyond it, inasmuch as it takes the cosmos to be a living theatre rather than a two-dimensional picture, and 'expresses' it rather than only 'painting' it.

However, the expressive function of ritual is not to be confused with a disorderly expression of individual feelings; rather it formalizes the right responses expected of the protagonist of the ritual, quite apart from its subsequent and variable (cathartic) effects.[19] This aspect is evident in *Ceremony* when Tayo resists his urge to fight back Emo's violence: his abstention from violence is the correct response, the one needed for the re-establishment of order both in his life and in the community.

It is the creative and generative power of words that the novel, as a piece of ritual literature or literary ritual, calls into being, not only their denotative, representational aspects. When Spindler talks of 'anthropological' magical realism, he is still in the mimetic frame of mind as it uses the term 'anthropological' as a filtering lenses that only

ISBN 978-0-9566139-1-2

'objective' onlookers can accept. But is there any such thing as an 'objective' onlooker? Nonetheless, as I have been arguing, it is possible and fruitful to employ anthropological and ethnographic methods and concepts when reading a piece of fiction.

The 'emic' – 'etic' tension and the patterning of the novel

In ethnography there is a distinction between an 'emic' and an 'etic' perspective: the first is the 'insider's view', while the 'etic' perspective is defined 'as the outsiders' view.[20] In her major work, *Pueblo Indian Religion*[21], anthropologist Parsons chose a '*non-systematizing, non-essentializing, anti-reductionist approach*'[22], in her attempt to show how '*religion is felt by the insider as an integral part of his life.*'[23] Thus she often tried to bridge the 'emic' – 'etic' gap. However, as an 'outsider' and an anthropologist relying on 'informants', she inevitably gave 'etic' descriptions of rituals, such as that of the 'Scalp dance' reported in her '*Notes on ceremonialism at Laguna*', where she reports the various stages of the ceremony:

> *It lasted a day and a night. The scalp was brought out hanging from the end of a stick which the women held, passing it from one to another. Before the dance the scalp was taken around to the houses of the opi, as the killers were called, and food and other things were requisitioned for the scalp. The opi […] had to keep a piece of the victim, a piece of skin or something else […] wrapped around their feet until the end of the dance. For twelve days after the kill the opi might not have sexual intercourse. Scalps were kept in jars in a cave to the north.*[24]

On the other hand, Silko gives a brief 'emic' account of the reasons behind the scalp ritual in a poem contained in the novel, where she highlights the negative consequences of omitting the Scalp ceremony for the warriors and their community: *'[…] They had things they must do/ otherwise/ K'oo'ko would haunt their dreams/ with her great fangs and/ everything would be endangered. / Maybe the rain wouldn't come/ or the deer would go away […]*'[25]

While Parsons never questioned the validity of science as an interpretative frame, Silko offers insight into Pueblo non-scientific mentality when she states:

> *Standing deep within the natural world, the ancient Pueblo understood the thing as it was – the squash blossom, grasshopper, or rabbit itself could never be created by the human hand - Ancient Pueblo took the modest view that the thing itself (the landscape) could not be improved upon. The ancients did not presume to tamper with what had already been created. Thus realism, as we now recognize it in painting and sculpture, did not catch the imaginations of Pueblo people until recently. […] Connection with the spirit dimensions requires a figure or form that is all-inclusive. A lifelike rendering of an elk is too restrictive. Only the elk is itself. A realistic rendering of an elk would be only one particular elk anyway. The purpose of the hunt rituals and magic is to make contact with all the spirits of the elk.*[26]

Thus 'realism' is 'too restrictive' and art must be 'all-inclusive', in order to tap into the spiritual dimension, which is also key to understanding the novel. Now, in *Ceremony*, is precisely this tension that is dramatized in the character of Tayo, as he is both an outsider and an insider at the same time. First, as a half-white child, then as a non-conforming war veteran he is regarded with suspicion by his aunt and the villagers. At the same time he is initiated into the secrets of Laguna cosmology by his uncle Josiah and it is thanks to this childhood training that later the medicine man Betonie is able to connect with him. Rocky, Tayo's cousin who dies in War World II, rejects his native traditions as superstitions as he is eager to be part of the whites' society and 'make something' of himself. Tayo, on the other side, is in a luminal state as Night Swan first introduces him to the necessary dynamics of change, he goes on to discover how all change works within and for a pattern of meaning. His search for 'meaning' has to be seen '*not in terms of "information"*' but in terms of '*pattern recognition*' and '*configurational awareness.*'[27] In other words, he sees how all fits in a cosmic cycle, regulated by the motion of the sun[28] through the seasons and days, as well as the transition from fertility to drought and back. Thus the form of the novel is '*the form of the ritual: form is the arrangement of contents*.'[29]

As Tambiah has argued, rites are characterized, among other things, by '*sequencing rules*'.[30] A look at the order of appearance of the various myths and chants in the novel can help clarify its performative character: the first myth is a 'cosmogony' about 'Thought-woman' thinking a story and the narrator as her messenger. Then the rationale for the ceremony is stated as what 'fight[s] off illness and death, the type of ceremony is a 'cure'. The switch into prose is preceded by the word 'sunrise'. After we learn about Tayo's condition as a traumatized war veteran, it follows the myth on the origin of drought from a misunderstanding between 'Reed Woman' and 'Corn Woman'. Then the first medicine man performs a Scalp ceremony for Tayo, and through a poem the reader comes to know about the ritual duties warriors have to fulfill in order to purify themselves from having killed or touched

ISBN 978-0-9566139-1-2

dead enemies. The novel progresses with various flashbacks into Tayo's memories, until another myth about the loss of ritual duties towards the Indians' mother goddess and the ensuing dearth and drought is related. This specific myth about animals trying to bring back fertility and plenty on the earth is broken into several chunks and is interwoven throughout the novel with Tayo's steps towards recovery. After learning about modern 'witchery' in Emo's drinking rituals and rehearsals of the old days in the army, Betonie sings how Indian witchery made white people, pointing to the lie that holds Indians enthralled:

> *That is the trickery of witchcraft, […] they want us to believe all evil resides with white people,. Then we will look no further to see what is really happening. They want us to separate ourselves from white people, to be ignorant and helpless as we watch our own destruction. But white people are only tools that they witchery manipulates; and I tell you, we can deal with white people, with their machines and their beliefs […]*[31]

Another important myth that is interwoven with the prose sections from the centre of the novel onwards is about the deer hunter turned into a coyote by Coyote god. It is first told by Betonie during the healing ceremony and then it is completed by the Laguna elders when they re-include Tayo in the community: clearly the mythical figure is embodied in Tayo himself and his predicament is Tayo's, as the Coyote in Pueblo Indian mythology is the trickster who works havoc in order to bring about change and growth.

In the same way, all the myths told are connected to the narrative by analogy. This pattern is particularly clear in the encounter between Tayo, the hunter and his wife T'seh, as already pointed out. The novel ends with the expulsion of Emo from the village and the respective chant of victory over witchcraft. The last poem is an '*offering to the sunrise*', which ends the novel on a note of optimism and ritual completeness, as it is fully contained between two dawns.

The myths and the prose are not only interspersed with each other, they parallel each other. On one occasion, Tayo remembers the story of the Gambler magician who locks away the clouds and father Sun going to their rescue. Later, this myth is related in its chant form, while Tayo clearly takes on father Sun's role when he rescues his uncle's cattle, whose 'speckled white' hides are compared to the very clouds. As the novel progresses, Tayo's actions become always more attuned to Pueblo wisdom as he performs all the 'right' things, in spite of the fact that he does not always remember the exact words and phrases.

Ceremony asks its readers to join the ritual, and thus functions according to a 'participatory' linguistic code that according to Tambiah characterizes art, religion and magic, while the switch to a 'causal' code happens in the sciences. Both codes depend on these two different and '*socially constructed orientations to reality*,'[32] the symbolic and the scientific. Language in the novel is a tool for specific actions, it is pragmatic, not an abstract system actualized in specific utterances according to the Saussurian distinction between *langue* and *parole*. The various chanted parts of the healing ritual are accompanied by what Tambiah calls '*object manipulation*,'[33] as when Betonie draws a Pollen Boy while singing the relevant myth.

CONCLUSION

It is evident that a definition of 'magical realism' based on an 'objective' point of view only cannot do justice to the complexity and inner variety of this literary mode. I have attempted to show how in the case of one specific novel the interplay of 'emic and 'etic' perspectives and the notion of 'performance' can replace the mimetic bias fruitfully. This bias is in fact both aesthetic and epistemological, in the sense of being founded on a culturally-specific vision of the world, the rationalistic, which is historically bound and is, however, not as monolithic as it may seem, given that even in the stronghold of rationalism and science, the post-industrialized West, many 'ethnic' or 'non-scientific' views coexist with scientific explanations.

Nonetheless, this bias covers also a political point, in the way 'outsiders' relate to so-called 'magical realist' texts. I hope to have shown that 'magical realism' is a 'dynamic' literary mode that negotiates what Tayo calls 'transitions through all distances and time.'[34] Finally, I also hope to have fulfilled the task I had set for myself, that of an interdisciplinary experiment in anthropology of literature.

ISBN 978-0-9566139-1-2

REFERENCES

[1] Zamora, L. P., and Faris, W. B. *Magical realism: theory, history, community.* (1995) Durham and London: Duke University Press, pp. 3-6.
[2] Connell, L. "Discarding magic realism: modernism, anthropology, and critical practice," in *ARIEL*, Vol. 29, No. 2, April, (1998), pp.95-110.
[3] Ousby, I, ed., "Magic realism", in *The Cambridge guide to literature in English*, (1992) Cambridge: Cambridge University Press, quoted in Connell.
[4] Connell, L. See note ii, p. 6.
[5] Lèvi-Strauss, C. *The savage mind* (1972) London: Weidenfield and Nicholson, p. 13.
[6] Connell, L. See note ii, p. 7.
[7] Tambiah, S. J. 'Form and meaning of magical acts', in *Culture, thought and social action.* (1985) Cambridge, Massachussetts, and London, England: Harvard University Press, p.79
[8] Tambiah, S. J. 'A performative approach to ritual', in *Culture, thought and social action*, p. 130.
[9] Tambiah, S. J. *Magic, science, religion and the scope of rationality* (1990) Cambridge: Cambridge University Press, p. 82.
[10] Silko, L. M. *Ceremony.* (1977) New York: Penguin Books, p. 100.
[11] Silko, L. M. *Ceremony*, p. 246.
[12] In Zamora and Faris, p. 409.
[13] Silko, L. M. *Ceremony*, p. 121.
[14] Silko, L. M. *Ceremony*, p. 126.
[15] Ibidem.
[16] Tambiah, S. J. 'The magical power of words', in *Culture, thought and social action*, p. 43.
[17] Ibidem, p. 48.
[18] Spindler, W. 'Magic realism: a typology', *Forum for Modern Language Studies.* (1993) Vol. xxxix No. 1, downloaded from http://fmls.oxfordjournals.org at University of Glasgow on April 17, 2010.
[19] Tambiah, S. J. 'A performative approach to ritual', in *Culture, thought and social action*, pp. 132-133.
[20] Wolcott, H. F. *Ethnography, a way of seeing.* (1999) Walnut Creek, CA: Altamira Press, p.137.
[21] Parsons, E. C. *Pueblo Indian religion* Vol. 1 and 2 (1996) Lincoln and London: University of Nebraska Press.
[22] Turner Strong, P. 'Introduction' to *Pueblo Indian religion*, Volume 1, p. xviii.
[23] Parsons, E. C. 'Preface' to *Pueblo Indian religion*, p. xxxiii.
[24] Parsons, E. C. 'Notes on ceremonialism at Laguna'. *Anthropological Papers of the American Museum of Natural History* vol. 19, 4, p. 122, downloaded from http://hdl.handle.net/2246/143 on April 20, 2010.
[25] Silko, L. M. *Ceremony*, p. 37.
[26] Silko, L. M. 'Interior and exterior landscapes: the Pueblo migration stories', in *Antaeus*, no. 57 (Autumn 1986) p. 2.
[27] Tambiah, S. J. 'A performative approach to ritual', in *Culture, thought and social action*, p. 139.
[28] For an informative article about this topic see Swan, E. 'Healing via the sunwise cycle in Silko's "*Ceremony*"', *American Indian Quarterly*, Vol. 12, No. 2 (Autumn 1988): University of Nebraska Press, pp. 313-328.
[29] Tambiah, S. J. 'A performative approach to ritual', in *Culture, thought and social action*, p. 143.
[30] Tambiah, S. J. Ibidem, p. 144.
[31] Silko, L. M. *Ceremony*, p. 132.
[32] Cunningham, G. *Religion and magic, approaches and theories* (1999) Edinburgh: Edinburgh University Press, p. 71.
[33] Tambiah, S. J. 'Form and meaning of magical acts', in *Culture, thought and social action*, p. 60.
[34] Silko, L. M. *Ceremony*, p. 246.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

REDUCING EMERGENCY DEPARTMENT CROWDING: BLURRING THE LINES BETWEEN HEALTHCARE, ENGINEERING AND MANAGEMENT

Kathleen BENIUK

PhD Student, Engineering Design Centre, University of Cambridge,
Cambridge, United Kingdom
kb424@cam.ac.uk

Keywords: Healthcare, Engineering, Management, Emergency Department, Crowding, Lean Six Sigma

ABSTRACT

Emergency department crowding (EDC) is a problem that many hospitals all over the world face and one, which can prove quite deadly. Whenever patient demand exceeds resource availability, EDC is bound to occur, resulting in adverse consequences including: ambulance diversions, increased waiting times, the delayed provision of crucial medical and nursing care, and increased patient mortality. Although EDC has been identified as a major concern in hospitals all over the world, there is little consensus on the underlying causes and even less agreement on solutions.

The objective of this study is to illustrate how an interdisciplinary approach, combining qualitative and quantitative tools drawn from healthcare, management and systems engineering, can effectively address a complex healthcare issue. This study will focus on the optimization of patient in, through, and outflow, with the aim of minimizing emergency department crowding. The process improvement technique Lean Thinking is used as the methodological framework for this study. This framework is comprised of five phases: define, measure, analyze, improve and control. Tools used include: a root cause analysis technique called a Current Reality Tree, a Delphi Study and semi-structured interviews to define and validate EDC, process mapping to highlight key and wasteful processes, simulation modeling to test possible solutions, an updated Dashboard system that alerts staff to the presence of crowding and the development of crowding action plans that will enable healthcare providers to alter their service provision quickly and efficiently to minimize the negative effects of crowding. Development and implementation of these improvements will require effective collaboration between healthcare providers, managers and engineers. This study illustrates how an interdisciplinary approach, combining tools and techniques from engineering, management and healthcare, can successfully contribute to improved emergency medical care.

INTRODUCTION

Emergency department crowding (EDC) is a problem that many hospitals all over the world face and one, which can prove quite deadly. Whenever patient demand exceeds resource availability, crowding is bound to occur. *'According to a 2002 national US survey, more than 90% of large hospitals report EDs operating "at" or "over" capacity'*,[1] a situation which can result in a number of adverse consequences including: ambulance diversions, increased waiting times, decreased quality of care, and the delayed provision of crucial medical and nursing care. In Australia, it is estimated that as many people die as a result of EDC as do in road accidents![2]

ISBN 978-0-9566139-1-2

Terminology

Currently there is no consistent term for the situation when patient demand exceeds the resource availability, when an emergency department is operating at or above capacity. *'A survey of the literature reveals that (two) terms, "crowding" and "overcrowding" are commonly used for this purpose, with most authors simply choosing one or the other of the (two) terms without explaining why they prefer it'.* [3] It seems logical that for clarity and consistency one term needs to be chosen. *'The term "overcrowding" suggests a more extreme situation and has a stronger negative connotation than "crowding"',* [3] implying that there exists a spectrum of crowding; from not crowded, crowding, crowded to a threshold when all of a sudden the ED is 'overcrowded' and therefore unmanageable. I disagree with this reasoning. There are many issues associated with a 'crowded' ED, which pose, *'significant moral risks to patients',* [3] and it is for this reason that I will be using the term ED 'crowding' thus forth.

Background Information

In all the literature reviewed, the earliest mention of EDC was a report in the journal '*Medical Care*' published in 1967 on non-urgent patients in the ED.[4] In the 1980's there were a few articles published discussing crowding. It wasn't until the early 1990's that ED Crowding became an identified research focus and since 2000 there has been an explosion of literature discussing the issue.[5] Research into crowding should be a top priority.

> *Crowding increases the risk of harmful medical errors in a variety of ways. In a crowded ED, errors may occur as a result of hurried treatment decisions with limited information, of delayed or poorly organized transfer of information from one clinician to another, or of failure to reexamine a patient or to reevaluate a previous physician's provisional diagnosis or treatment plan.*[3]

A study published by Derlet in 2001 reported that EDC in US Emergency Departments resulted in delays in diagnosis and treatment ultimately leading to poorer health outcome[6] and a report released by the Joint Commission on Accreditation of Healthcare Organizations stated that EDC was a factor in almost a third of all patient deaths examined.[7]

Crowding is undeniably contributing to a decrease in quality care, but what can be done?

Objectives

The objective of this study is to optimize patient flow in, through and out of emergency departments with the aim of reducing crowding. Due to the *'critically important ramifications of the quality of emergency health care rendered to patients presenting to a hospital seeking emergency services'*[8] I believe that the best approach is to *'look at the issue first from a systems perspective to identify root causes; and then propose and test potential solutions without perturbing real patients or real staff'*.[8] Because EDC is such a complex problem, it will require a multidisciplinary approach and the collaboration of a multidisciplinary team. This study will use a mixed-method approach, combining qualitative and quantitative tools from management and systems engineering to formulate and test healthcare solutions. This study will take place at the Cambridge University Hospital NHS Foundation Trust, Addenbrooke's Accident and Emergency Department (A&E), in Cambridge, England.

METHODOLOGY

Lean Six Sigma Framework

The NHS Institute for Innovation and Improvement defines Lean as *'an improvement approach to improve flow and eliminate waste'*.[10] Lean was first introduced by Toyota in the 1930's through the development of the Toyota Production System.[9] Toyota was attempting to improve on Henry Ford's process line manufacturing methods by shifting *'the focus of the manufacturing engineer from individual machines and their utilization, to the flow of the product through the total process'*.[9] This method of process improvement is now known as Lean Manufacturing or Lean Thinking. In 1981 Motorola created their own business management strategy called Six Sigma which was based on the Toyota method.[11] Six Sigma uses a strict project methodology and statistical analysis to identify and eliminate defects. Today, these two methods are often combined, referred to as Lean Six Sigma, and the technique has been applied to many industries including healthcare. What makes this technique so successful in different industries is its use of *'new concepts, tools and methods'* [10] that *'address workplace organisation, standardisation, visual control and elimination of non-value added steps'*[10] to *'improve flow, eliminate waste and exceed customer expectations'*. [10]

ISBN 978-0-9566139-1-2

This study uses the Six Sigma five-phase DMAIC methodological framework (define, measure, analyze, improve and control) with Lean and engineering process improvement tools to identify, assess, and reduce emergency department crowding.

Define

The aim of the define phase is to identify the project goals and the intended improvements. Common Lean tools used in this stage include brainstorming, high level process mapping, including current reality trees, and stakeholder analysis.[12] When planning a Lean Six Sigma project, Pyzdek in his book *'The Six Sigma Project Planner: A Step-by-Step Guide to Leading a Six Sigma Project Through DMAIC'* identifies questions that should be answered at each phase of the DMAIC framework. [13] Below are five questions to be answered in the define phase.

What is the business case for the project?
By reducing crowding we can improve patient flow, reduce unnecessary processes, reduce patient's total length of stay, and improve the overall quality of care while reducing the overall financial costs of providing care.

Who is the customer?
Patients, hospital staff, the NHS.

What is the current state map?
To develop an accurate picture of the current state and issues associated with EDC, I used a Lean tool called a Current Reality Tree (see Tool #1 below). This tool helped identify the link between effects of EDC to their underlying root causes.

What is the scope of the project?
The scope of this project is to answer the following questions: what is EDC, what are the underlying causes and what is and can be done to minimize EDC given the current available resources and expected patient demands?

What are the deliverables?
The deliverables I will develop as a result of this project are a quantifiable definition of EDC, an integrated software tool that will allow staff to instantly detect when and why the ED is crowded and situation specific action plans to reduce EDC when it is occurring.

Tool #1 - Current Reality Tree

A current reality tree is an efficient method of analyzing a system. The aim of this tool is to pictorially demonstrate the relationships between the undesirable effects of an issue and its identified root causes,[14] providing a focus for further research. The first step to building a current reality tree is to create a list of undesired effects you wish to address. Focusing on one effect at a time, generate a list of possible contributing factors, then for each contributing factor list their possible contributing factors. This cycle continues until you have clearly identified the root cause of each effect.[14] Root causes can be sub-divided into those that can be controlled and those that cannot. For example, it is difficult to accurately predict and control the number of patients that will arrive at the ED each day. Alternatively, it is possible to control where certain beds or supplies are located in the department. The challenges faced when generating a current reality tree are clearly defining each undesirable effect, identifying the true root causes, and understanding how the complex processes are interconnected.

This study used information from published literature to identify the undesirable effects of ED crowding and their root causes.

Measure

After the define phase clearly identifies the project goals and intended improvements, the project progresses to the measure phase. The aim of the measure phase is to evaluate the *'source of a problem as precisely as possible by building a factual understanding of existing process conditions'*.[15] This stage often involves data collection and analysis, and common Lean tools include flow charts, histograms and check sheets.[12] Below are five questions that should be addressed in the measure phase.[13]

ISBN 978-0-9566139-1-2

What are the key metrics for this business process?
'A fundamental difficulty in studying ED crowding is the lack of a universally accepted definition' [16] of what exactly ED crowding is and the key metrics required to measure it. *'It may be obvious to the average person when an ED is overcrowded, but definitions based on precise wait times, or quantitative delays in actual ED care are lacking'*.[17] To develop a universally accepted definition of EDC I used a formal consensus technique called a Delphi Study to identify measures that would function as an internationally recognized clinical definition for Emergency Department crowding (see Tool #2 below).

Are the metrics valid and reliable?
To validate if the measures defined in the Delphi Study were a reliable method of defining and measuring EDC, I performed a series of semi-structured interviews of ED nurses, doctors, paramedics, and managers at Addenbrooke's A&E (see Tool #3 below).

Do we have adequate data on the process?
ED patient data is collected at Addenbrooke's. This includes the patients' medical history, treatment plans and test results. The adequacy and quality of this data will need to be evaluated to ensure the results of this project are accurate. The data quality is a potential area of concern as most is manually entered by ED staff who, when the ED is busy, must delay data entry to prioritize patient care. It is for this reason that I am apprehensive about the accuracy of this data entry.

What is the baseline?
Historical data collected in the ED will provide the baseline for any future evaluation.

How will I measure project success?
The project success can be measured in one of two ways. To measure whether or not there have been any significant improvements in reducing the occurrence of EDC, the rate of measured EDC in data obtained after the intervention can be compared to historical data. Another way of measuring success would be to contrast ED staff experiences of crowding, as indicated in the semi-structured interviews, before and after the intervention.

Tool #2 - Delphi Study
'The Delphi process takes its name from the Delphic oracle's skills of interpretation and foresight'.[18] A Delphi Study is a formal consensus technique that uses *'a series of sequential questionnaires or "rounds", interspersed by controlled feedback, that seeks to gain the most reliable consensus of opinion of a group of experts'*.[19-21] This method has been used widely in health research. [19, 21] *'The Delphi technique and other consensus development methods should not be viewed as a scientific method for creating new knowledge, but rather as processes for making the best use of available information, be that scientific data or the collective wisdom of participants'*.[19] The, '*main advantage of the Delphi is reported to be the achievement of consensus in a given area of uncertainty or lack of empirical evidence'*.[19] This method is also able to overcome the disadvantages normally found in group or committee decision making,[18, 22] which *'are recognized to be prone to domination by powerful individuals, the biasing effects of personality traits, seniority and the fact that only one person can speak at a time'*.[19]

The aim of this study was to develop an internationally recognized clinical definition for Emergency Department crowding.

Delphi study '*participants should be experts who reflect current knowledge and perceptions, yet are relatively impartial to the findings*'[19], they should be, '*drawn from varied backgrounds in order to guarantee a wide base of knowledge*'.[19] In this Delphi Study, participants were chosen in one of three ways. The first group of experts invited to participate were professional contacts of my study collaborator, Dr. Adrian Boyle, or myself. The second and third groups of experts were selected because of their active involvement in publishing journal articles pertinent to this field. Each expert was contacted via email, given information about the study and the predicted time requirements, and invited to participate.

This study was completed entirely online using the online survey tool Survey Monkey. The benefit of using an online survey tool over mailed surveys, phone or personal interviews was that, provided each participant had access to the internet, they were able to quickly and securely access and submit confidential responses. It also made providing feedback to participant's inquiries quick and efficient.

ISBN 978-0-9566139-1-2

For this study, a three round Delphi study was performed using methods drawn from previous Delphi studies found in the literature.[18-21, 23-27] The purpose of the first round of the Delphi study is to identify broad issues that will be further assessed in the succeeding rounds. To do this, *'the first round questionnaire is usually unstructured and seeks an open response'*.[19] In this study, participants were asked to list all possible defining characteristics that should be included in the new definition of ED crowding. A defining characteristic was defined as *'one of a number of essential features by which a crowded ED can be recognized'*.[20, 26] For each defining characteristics, respondents were asked to submit an operational definition which was defined as a '*set of directives, activities or procedures that specify how to measure, observe or record the defining characteristic*'.[20, 26] These statements are collected, categorized, and sent back to participants for further evaluation. *'The second and subsequent rounds are more specific, with the questionnaires seeking quantification of earlier findings, usually through rating or ranking techniques. Because the researcher feeds back results from previous rounds, there tends to be convergence to a consensus of opinion'*.[19] Responses from round two and round three were evaluated against a predetermined percentage level for inclusion in the definition. For round two the agreement rate to maintain a defining characteristic or operational definition was set at 50% agreement, and in round three, an agreement rate was set at 70%. This level was consistent with what had been found in the literature.[20, 28, 29] Upon completion of round three, it was decided that a consensus had been reached and for this study and no further rounds were required. The EDC definition developed consists of quantifiable measures that can be used to identify precisely when and why crowding is occurring.

Tool #3 - Semi-structured Interviews

'Interviews are the most commonly used qualitative technique in health care settings'.[30] There are three types of interviews: structured, semi-structured and depth interviews. Structured interviews have a structured standardized questionnaire with fixed answer choices, for example yes or no. Semi-structured interviews are '*conducted on a loose structure consisting of open-ended questions that define the area to be explored*'.[30] Depth interviews are the least structured and focus on one or two issues in great detail asking broad open-ended questions.[30] I performed a series of semi-structured interviews of ED nurses, doctors, paramedics, and managers at Addenbrooke's Hospital Accident & Emergency. The aim of this study was to validate the international definition of EDC developed in the Delphi Study, confirming it is a reliable method of measuring EDC.

The questions asked of participants were both open and closed. Closed questions seek a specific answer, for example 'does the department become crowded, yes or no?' Open questions require the participant to elaborate upon their answer with detail, for example 'when is the department likely to be crowded?' The phrasing of the questions is of the utmost importance. The questions must guarantee a predictable and reproducible effect. Each participant must be able to understand the questions, specific terminology must be defined, and the questions must be phrased in a way that they do not appear to be leading to a specific answer. To ensure the interviewee accurately captures the responses, it is advisable to tape record each interview and transcribe the responses afterwards. Interview length will vary but for my study, each interview lasted no longer than 20 minutes. The interview responses were examined by means of thematic analysis; the data was analyzed and recurrent themes identified. These recurrent themes were contrasted against the measures identified in the EDC definition to evaluate the validity of the definition.

Analyze

The aim of the analyze phase is to *'eliminate the gap between the current performance and the desired goal'*.[15] Common Lean tools used in this stage include cause and effect diagrams, failure modes and effects analysis and pareto analysis.[12,15] Below are five questions to address in the analyze phase.[13]

What is the current state?

To determine the current state and identify key and wasteful processes, I will develop a process map known as value stream mapping (see Tool #4 below). This tool is commonly used in both management and systems engineering.

Is the current state as good as the process can do?

To evaluate the behavior of the current system I will use a tool from systems engineering, simulation modeling (see Tool #5 below). By simulating the ED as it is now and testing it under a number of different conditions, I will be able to identify if and how the system can be improved.

ISBN 978-0-9566139-1-2

Who will help make the changes?
I will work with ED staff; nurses, doctors and managers, to facilitate the implementation of the changes in ED service provision.

What are the resource requirements?
The changes in resource requirements will be identified in the simulation model and written into the crowding action plans (see Tool #6 below).

What could cause this change effort to fail?
'Research has shown that regardless of the location, instigating change is difficult, especially if it requires complex changes such as in a clinical practice, or an increase in collaboration between disciplines'.[31, 32] Achieving and sustaining *'successful implementation of health policy is a function of the relation between the nature of the evidence, the context in which the proposed change is to be implemented, and the facilitation approach adopted for the change'.*[33] Without the right balance between balance, context and facilitation there is little chance that this change will be adopted and institutionalized. By providing strong evidence that the intervention will minimize EDC, focusing the intervention on the specific contextual requirements of the ED and by empowering staff to take ownership of facilitating the change I believe that this intervention can be successful.

Tool #4 – Process Mapping
Process mapping is an effective way of modeling the flow of people, products or services within a system. Process mapping was first formally introduced to the world by one of the founders of Industrial Engineering, Frank Gilbreth, in his 1921 presentation entitled '*Process and Flow Process Charts*' at the American Society of Mechanical Engineers (ASME) Annual Meeting.[34] Since then process mapping has become very popular in a whole range of industries, including healthcare.

This particular type of process mapping used in this study is called Value Stream Mapping, which is quite common tool in Lean Thinking. *'A value stream is defined as all the actions (both value added and non-value added) required to bring a specific product, service or a combination of products and services, to a customer'.*[8] All process maps use a set of standardized symbols, that were initially established the ASME in 1947, to represent various aspects of the system.[34] The first step to developing a process map is to identify the various products or services that are in operation and to select one to focus on, thus setting the boundaries of the study. Then you would gather actual process data to generate a current state map for the selected product or service. The current state map will draw into light a number of non-value added processes which can be altered or eliminated in the development of a future state map.[8] In this study the process map will be used as a template for the simulation model.

Tool #5 - Simulation Modeling
A computer simulation is a model, an imitation of a complex real world system. The benefit of simulation is that at low cost and low risk, you are able to better understand the system and accurately predict how it would behave under different conditions. This can enable you to make better decisions. Building a simulation model starts with a scope of the problem. This is useful to determine the boundaries of the simulation, the limitations or restrictions and the expected outcome. It is essential to identify what part or object you will be tracking through the simulation. This could be people, products or information. Next you need to identify the resources involved in the process. This can include people, parts, and machines, and data about their resources availability, reliability and ability needs to be collected. Once the resources have been identified, you need to identify the events or processes that are taking place. Information about these processes that is needed includes processing times, flow and methods. This information can come from the process map. This information is all input into the simulation modeling software you have selected. There are a number of different simulation software programs, each with their own strengths. Examples of these include Arena, Simio, Simulink, and Simul8. Once you have been able to verify the simulation model is a valid and credible representation of the system as it is, you are able to modify the simulation to illustrate how the system would behave in different situations.

For this project, I will create a model of the emergency department at Addenbrooke's Hospital. This model will be developed with the Simio simulation modeling software.[35] I chose Simio because it is one of the newest simulation programs, it has a clear and intuitive user interface, and easily converts simulations into 3D animated

ISBN 978-0-9566139-1-2

models. The simulation will be built with actual (anonymous) historical patient data collected by Addenbrooke's Operations Centre. This data required includes patient arrival rates, patient acuity, staffing schedules and bed availability. Simulating the ED as it currently operates will identify whether crowding is actually occurring and quantify when and how often. I will also be able to isolate specific causes and effects of crowding and identify any recognizable patterns. By then modifying the simulation I will be able to observe how the system behavior changes. By testing a number of situational variations against the current system I will be able to identify specific service modifications that minimize crowding. These results will support the implementation of situation specific crowding action plans to alleviate the effects of EDC.

Improve

The aim of the improve phase is to develop a solution and implementation plan and evaluate its effects.[12, 36] Common Lean tools used in this stage include tree diagrams, check sheets and flow charts.[12] Below are three questions that should be answered in the improve stage.[13]

What is the work breakdown structure?
The work breakdown structure will be outlined in crowding action plans (see Tool #6 below).

Future state map?
In the ED it is impossible to identify the optimal state map because the system demands are always changing. A number of ideal future service designs corresponding with various common situations will be identified in the simulation model and translated into crowding action plans.

What specific activities are necessary to meet the project's goals?
The aim of this study is to optimize patient flow in, through and out of the ED to reduce crowding. To achieve this there will need to be a positive culture of change and improvement. Specifically there will need to be support from management, there will need to be IT support to both collect and analyze patient data, clinicians and ED nurses will need to be flexible and open to altering their methods of care. Finally patients will need to be educated and empowered to expect not only effective but efficient care.

Tool #6 – Crowding Action Plans

The action plans are a product of the simulation model. Each action plan will target specific situations that have been identified as negatively impacting patient flow and resulting in crowding of the department. For example, if there was a sudden increase in the inflow of patients increasing ED occupancy, the action plan will outline advisable actions or changes in service provision that, if implemented, will alleviate the symptoms of crowding. The crowding action plans will work in conjunction with the Dashboard system (see Tool #7). Crowding action plans will target each of the measurable identifiers as outlined in the EDC definition.

Control

The aim of the control phase is to verify the process has changed and to control and maintain the new system.[12,36] This phase ensures that any changes are maintained and institutionalized into the culture of the department and encourages continual assessment and process improvement. Common Lean tools used in this stage include dashboards, control charts and affinity diagrams.[12] Below are three questions to be answered in the control phase.[13]

What types of progress reports should I create?
Progress will be tracked using a Lean tool called a Dashboard (see Tool #7 below). This tool will alert staff when EDC is occurring and reference the specific crowding action plan they should institute to minimize crowding.

How will I assure that the business goals of the project were achieved and maintained?
The dashboard system will rely on up to date patient data that, when compared to historical data, will give an accurate indication of the improvement in patient flow that has been made as a result of the crowding action plans.

ISBN 978-0-9566139-1-2

How will I keep the gains made?
To maintain the gains achieved by this project it will require the continued support of ED staff and management.

Tool #7 - Dashboard
'A dashboard is a tool used for collecting and reporting information' [37] which *'displays clear, measureable and valid metrics for each objective, targets for each metric, and the status of each metric'*.[37] Often the metrics are color coded so that their status can easily be identified. This tool is easy to understand and provides real-time information on performance that can enable better, more informed and timely decision-making.

A Dashboard system is already in place at the Addenbrooke's A&E. This screen displays specific patient data including total department patient occupancy, waiting times and total length of stay. My proposal is to modify this dashboard to also include a metric that alerts staff when crowding is occurring and references the specific crowding action plan they should institute to minimize crowding. This metric would also provide an accurate measurement of ED crowding and its specific causes.

CONCLUSION

The objective of this study was to use a multidisciplinary approach to optimize the flow of patients in, through and out of emergency departments. This study has, thus far, successfully incorporated process improvement methods and tools from management and systems engineering to confront healthcare challenges and improve emergency care.

REFERENCES

[1] Trzeciak, S. and E.P. Rivers. *Emergency Department Overcrowding in the United States: An Emerging Threat to Patient Safety and Public Health.* (2003) Emerg Med J. vol 20, no. 5: 402-5.
[2] Johnston, M. "*Hundreds' Die Because of Hospital Crowding*', in *The New Zealand Herald.* 8 September 2008: Auckland.
[3] Moskop, J.C. et al. *Emergency Department Crowding, Part 1--Concept, Causes, and Moral Consequences.* (2009) Ann Emerg Med. vol 53, no. 5: 605-11.
[4] Kirkpatrick, J.R. and J.T. Leon. *The Non-Urgent Patient on the Emergency Floor.* (1967) Medical Care. vol 5, no. 1: 19-24.
[5] Moskop, J.C. et al. *Emergency Department Crowding, Part 2--Barriers to Reform and Strategies to Overcome Them.* (2009) Ann Emerg Med. vol 53, no. 5: 612-7.
[6] Derlet, R., Richards, J., & Kravitz, R. *Frequent Overcrowding in U.S. Emergency Departments.* (2001) Academic Emergency Medicine. vol 8, no. 2: 151-5.
[7] Joint Commission on Accreditation of Healthcare Organizations. '*Sentinel Event Alert.*'. (2002), from http://www.jcaho.org
[8] Koelling, C.P. et al., '*Value Stream Mapping in the Emergency Department*'. Grado Department of Industrial and Systems Engineering, Virginia Tech. Blacksburg, VA.
[9] Lean Enterprise Institute. '*A Brief History of Lean*'. (2009), Retrieved 17 May, 2010, from http://www.lean.org/whatslean/History.cfm
[10] National Institute for Innovation and Improvement. '*Lean Thinking*'. (2006), Retrieved 17 May, 2010, from http://www.institute.nhs.uk/quality_and_value/lean_thinking/lean_thinking.html
[11] Tennant, G. *Six Sigma : Spc and Tqm in Manufacturing and Services.* (2001) Aldershot, England; Burlington, VT. Gower.
[12] Davies, J. *Lean Six Sigma Process Improvement: Lecture 1a.* (2010) Judge Business School, University of Cambridge. (Unpublished Work)
[13] Pyzdek, T. *The Six Sigma Project Planner a Step-by-Step Guide to Leading a Six Sigma Project through Dmaic.* (2003) New York; London. McGraw-Hill.
[14] Dirgo, R. *Look Forward Beyond Lean and Six Sigma a Self-Perpetuating Enterprise Improvement Method.* (2006) Fort Lauderdale, Fla. J. Ross Pub. : Co-published with the Institute of Aerospace Excellence.
[15] Davies, J. *Lean Six Sigma Process Improvement: Lecture 2a.* (2010) Judge Business School, University of Cambridge. (Unpublished Work)
[16] Bernstein, S.L. and B.R. Asplin. *Emergency Department Crowding: Old Problem, New Solutions.* (2006) Emerg Med Clin North Am. vol 24, no. 4: 821-37.
[17] Derlet, R.W. and J.R. Richards. *Overcrowding in the Nation's Emergency Departments: Complex Causes and Disturbing Effects.* (2000) Ann Emerg Med. vol 35, no. 1: 63-8.
[18] Webster-Harrison, P., A. White, and J. Rae. *Acupuncture for Tennis Elbow: An E-Mail Consensus Study to Define a Standardised Treatment in a Gps' Surgery.* (2002) Acupunct Med. vol 20, no. 4: 181-5.
[19] Powell, C. *The Delphi Technique: Myths and Realities.* (2003) J Adv Nurs. vol 41, no. 4: 376-82.

ISBN 978-0-9566139-1-2

[20] Ferguson, N.D. et al. *Development of a Clinical Definition for Acute Respiratory Distress Syndrome Using the Delphi Technique*. (2005) J Crit Care. vol 20, no. 2: 147-54.
[21] Yousefi-Nooraie, R. et al. *Teaching Evidence-Based Practice: The Teachers Consider the Content*. (2007) J Eval Clin Pract. vol 13, no. 4: 569-75.
[22] Surowiecki, J. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few*. (2005) United States. First Anchor Books: Random House, Inc.
[23] Beattie, E. and K. Mackway-Jones. *A Delphi Study to Identify Performance Indicators for Emergency Medicine*. (2004) Emerg Med J. vol 21, no. 1: 47-50.
[24] Greenhalgh, T. and T. Wengraf. *Collecting Stories: Is It Research? Is It Good Research? Preliminary Guidance Based on a Delphi Study*. (2008) Med Educ. vol 42, no. 3: 242-7.
[25] Fleuren, M., K. Wiefferink, and T. Paulussen. *Determinants of Innovation within Health Care Organizations: Literature Review and Delphi Study*. (2004) International Journal for Quality in Health Care. vol 16, no. 2: 107-23.
[26] Grant, J.S. and M.R. Kinney. *Using the Delphi Technique to Examine the Content Validity of Nursing Diagnoses*. (1992) Nurs Diagn. vol 3, no. 1: 12-22.
[27] Pines, J.M. *Moving Closer to an Operational Definition for Ed Crowding*. (2007) Academic Emergency Medicine. vol 14, no. 4: 382-3.
[28] Roberts, F.E. *Consensus among Physiotherapists in the United Kingdom on the Use of Normal Saline Instillation Prior to Endotracheal Suction: A Delphi Study*. (2009) Physiother Can. vol 61, no. 2: 107-15.
[29] Suwaratchai, P. et al. *Utilize the Modified Delphi Technique to Develop Trauma Care Indicators*. (2008) J Med Assoc Thai. vol 91, no. 1: 99-103.
[30] Pope, C. and N. Mays. *Qualitative Research in Health Care*. (2006) Malden, Mass. Blackwell Pub./BMJ Books.
[31] Grol, R. and J. Grimshaw. *From Best Evidence to Best Practice: Effective Implementation of Change in Patients' Care*. (2003) Lancet. vol 362, no. 9391: 1225-30.
[32] Beniuk, K., '*First Year Report*'. (2009) University of Cambridge, Department of Engineering. Cambridge.
[33] Jun, G.T., K. Beniuk, and P.J. Clarkson. *What Service Designers Can Learn from Policy Designers* (2010) Touchpoint - Service Design Journal. vol 2, no. 1: 102.
[34] Graham, B., '*The Root of the Business Process Mapping*'. (2008)
[35] Simio LLC. '*Simio: Forward Thinking*'. (2009), Retrieved 2010, 31 May, from http://www.simio.biz/
[36] Davies, J. *Lean Six Sigma Process Improvement: Lecture 3a*. (2010) Judge Business School, University of Cambridge. (Unpublished Work)
[37] i Six Sigma. '*Quality Dictionary and Glossary: Dashboard*'. (2000), Retrieved 24 February, 2010, from http://www.isixsigma.com/dictionary/Dashboard-218.htm

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

THE THEOLOGY OF SPATIAL EQUALITY AND SOCIAL INTEGRATION

Kevin J. BROWN

PhD Student, University of Glasgow, Scotland
kjbrown2@anderson.edu

Keywords: Segregation, Ethics, Social Integration, Theology, Community

ABSTRACT

For an advanced democracy, America is one of the most segregated countries in the world. In the midst of this, efforts towards integrating communities and persons has had mixed results, and persuasive arguments have even been advanced suggesting an economic benefit to allowing for inert, segregated communities. However, the ambiguity over appeals for spatial equality and integration has stretched the criteria, and thus the voices relevant to this important topic, creating a more comprehensive, interdisciplinary approach. In this paper, my intention is to continue the dialogue by presenting an ethical appraisal of existing work in this area. More specifically, I aim to do this by providing a theological perspective related to the concept of integration, and explicate its value for society today. Normative theological inquiry provides an egalitarian voice to our otherwise market dominated perspectives in this field, and further offers a picture of integrated social arrangements through a clear biblical trajectory of transcending racial, socio-economic, and cultural boundaries in the new covenant.

INTRODUCTION

This paper aims to present the problem of US housing segregation and the difficulty in prescribing a definitive solution (i.e., mixed communities or social integration). In addition to casting an ethical light on this important subject, I specifically aim to present a theological approach towards the issue of social integration. I submit that a theological voice offers a picture of integrated social arrangements that transcends the otherwise narrow perspectives that have dominated this field of inquiry. Further, it offers an ideal of morality less susceptible to human self-interest (regards towards individual or corporate maximizations of pleasure and minimizations of pain), and is more anchored to deontological principles of egalitarianism, equality, and justice as well as to the teleological ends of welfare, well-being, and community. In addition to a theological perspective offering a unique lens by which to appraise, interpret, and present the issue of segregation and the idea of mixed-communities, such a perspective assists us in reframing our definition of prosperity.

Divided We Stand

Of all of the terms used to describe America, both disparaging and complimentary, there is perhaps none that more accurately describes the 50 states than this: segregated. The statistical measure most often referenced when determining segregation is the dissimilarity index, which represents the degree to which blacks and whites are evenly spread among neighborhoods in a city. [1] According to Massey and Denton (1993), scores above 60 represent high levels of segregation, and scores below 30 represent low levels of segregation.[2] Ironically, the years following the American Civil War, a war fueled by the issue of slavery and black rights, were the years that represented some of the lowest numbers of segregation

ISBN 978-0-9566139-1-2

our country would know. Dissimilarity Indices collected from a group of major metropolitan areas in both northern and southern US regions show racial dissimilarity between 29 and 46 around the time of the civil war. Ranges of 38 to 59 are provided circa 1910, with a massive increase in dissimilarity in the range of 81 to 89 in 1940.[3] In 1968, US President Lyndon Johnson created a committee to address the violence that was erupting through rioting in the nation's distinctly African-American ghettos. Among other conclusions, the committee reported that the US was '*moving toward two societies, one black, one white—separate and unequal.*'[4]

After World War II, structural patterns of black-white separation began to emerge that would not only increase the trend of segregation, but ensure that this segmented living structure would become a blueprint for the future of US residential housing. With a historical catalogue of incidences such as discrimination in Federal mortgage guarantees, exclusionary zoning policies, and other forms of structural and individual level discrimination, it comes as no surprise that in the 1970s, the pattern of a black core surrounded by a white ring defined the city to suburb relationship in the US.[5] By early 1990s, nearly one quarter of all US African Americans could be found in 10 US metropolitan areas.[6] Finally, at the turn of the century, the average white person in Metropolitan America lives in a neighborhood that is 80% white and only 7% black. Conversely, the average African American lives in a neighborhood that is 33% white and as high as 51% black.[7] This alarmingly unequal treatment of African Americans led Massey and Denton to conclude in their landmark study of US segregation that African Americans are '*unambiguously among the nation's most spatially isolated and geographically secluded people, suffering extreme segregation across multiple dimensions simultaneously.*'[8]

This conclusion, and the accompanying statistics to substantiate it, is troublesome. Even more troubling is the dissonance between current measures of segregation and the often cited US mantras of '*Liberty and Justice for All*' and '*United We Stand*'. Should either of these phrases really be housed under the shelter of US ideology when we are highly segmented in reality? This question has not gone unnoticed by legislators who have responded with efforts to integrate communities by attaching economic disincentive[9] to segregation in addition to the creative use of policy tools. Describing the underlying motivation for this policy, Sociologist John Logan writes: '*Neighborhood integration has remained a goal of public policy and popular opinion because it is seen as proof of the American ideal of equal opportunity.*'[10]

However, efforts towards reducing residential segregation and building integrated communities have been met with legitimate criticism. As egalitarian as social integration policy may appear in nature, measurements towards the social benefits of mixing have been difficult to substantiate in terms of fiscal and social efficiency. As Blasius et al (2007) summarizes, '*The adequacy of the evidence base to support this position has been the subject of spirited debate on both sides of the Atlantic.*'[11] When surveying a great portion of social integration skepticism (such as the aforementioned quote), the term 'adequacy' often serves as a euphemism for economic or utilitarian outcomes: ideologies famous for the 'ends justifying the means'. For example, one can look at the language that followed the Mt. Laurel exclusionary zoning trials throughout the 1980s and 1990s to see this principle at work. In Mt. Laurel, a seminal land usage case in the US, the New Jersey Supreme Court Justices ruled that exclusionary zoning was morally wrong and that provision had to be made for low to moderate income families to have opportunities to live within the municipality.[12] This rule was not taken well, and the white framers of the original exclusionary zoning policy fought this proposed arrangement:

> *The townships hired social scientists to churn out documents contending that poor people were truly happier with their own kind; that living in the suburbs would be bad for them and, besides, would mean an end to the suburban way of life—a variant, for land use, of the infamous argument that blacks were inferior and so would be better off attending segregated schools.*[13]

As this quote evidences, ethical/moral judgments were off the table and replaced with the utilitarian sentiment that everyone will be happier under the status quo. This brings up an important question: if a wide range of research has demonstrated integration to be sub-optimal, what is the criteria being employed to make this judgment? Further, if economic (financial justification) or utilitarian (justification through the maximization of gross happiness) criteria dominate the repertoire of evaluative measures, perhaps it is worth expanding our criteria to employ other tools across the evaluative

ISBN 978-0-9566139-1-2

landscape as it relates to our judgments on social integration. In the wake of a systematic international financial crisis, our present context is an ideal point in time to undertake this exercise.

Expanding Our Ethical Horizons

At the close of 2009, the otherwise self-proclaimed free-trade/free-market authors writing for *The Economist* took stock of the term 'progress' and its contemporary definition.[14] On the heels of the financial meltdown, one might expect a ready defense for the existing free-market structure.[15] After all, our contemporary situation still boasts the position of being the most prosperous century on historical record. An historical survey of past periods of economic, scientific and technological prosperity have no parallel when compared to our present context, and this alone is reason enough, it might be argued, for pause when considering the distribution of blame in the wake of the financial crisis.

However, in their article '*Onwards and Upwards*', this was not the inevitable conclusion reached.[16] Indeed, amidst centuries of increased efficiency in producing food, enlightenment science, industrial growth, technological innovation, and gains in overall wealth among both rich and poor nations, such 'material progress' has failed to deliver emotional satisfaction and overall happiness. Rather, they write, one of the more salient outcomes of our prosperity is the attrition of life around us: '*The forests are disappearing; the ice is melting; social bonds are crumbling; privacy is eroding; life is becoming a dismal slog in an ugly world.*'[17] Citing the philosopher Susan Neiman, *The Economist* suggests that our behavior should not be shaped by power, material wealth, etc., but rather by what is 'right' despite the inconveniences that accompany the pursuit of this ideal.[18] Aligning themselves with Neiman, they contend that this is a more appropriate measurement of progress and prosperity.

The point raised by *The Economist* is a valid one, and certainly worth our consideration in the wake of perhaps the worst international financial crisis our contemporary world has ever experienced. At this point, my aim is to explore the issue of segregation and the natural response of social integration utilizing an expanded criteria base that incorporates a thorough moral appraisal of this phenomenon. I argue that such an exercise is not simply helpful or supplementary, but necessary given the economic ambiguity associated with the efforts towards mixing communities as a response to significant levels of US segregation.

DISCUSSION

Mixed Communities

Some researchers contend that social interactions and market dynamics tend to produce segregated rather than integrated communities.[19] However, it would be a mistake to simply attribute the status quo of residential segregation to mere market forces. Black-white dissimilarity indices still remain highly uneven when controlling for differing levels of income.[20] While market interactions may widen the gap between races, residential segregation has also been explained as a function of policy choices—a gap that '*has been brought about and maintained by rules of a game that operates as a subtle apartheid.*'[21] Further, African American preferences for integration are historically much higher than whites. Massey and Denton's (1993) research has shown that the vast majority of African Americans express strong support for integration. On surveys, when asked about whether they favor 'desegregation, strict segregation, or something in-between' they have answered 'desegregation' in large numbers.[22]

Because segregation cannot simply be attributed to market forces or market preferences, the inequality of the situation is heightened considering segregation in and of itself creates detrimental consequences for those inhabiting the segmented area. In fact, it has been described as the definitive source responsible for the perpetuation of black poverty in the US.[23] For example, Chenoa Flippen's work has demonstrated that segregation has a deleterious effect on housing values for black segregated areas.[24] Further, segregation has been positively associated with increased unemployment, poorer educational results, and neighborhood crime.[25]

ISBN 978-0-9566139-1-2

Yet the problem has not gone unaddressed. Political efforts towards integration have been attempted in numerous forms for several decades. Despite the government's use of various integration measures such as dilution, diversity, and dispersal[26], to classify such policy as being informed by evidence would be a mistake. Indeed, the enterprise of mixing communities rests on the belief that mixing is good and offers utility for those inhabiting the same space, particularly the disadvantaged. Hardman and Ioannides write:

> *The value of neighborhood interactions has attracted policymakers' attention and led to policy initiatives intended to take advantage of positive externalities associated with mixing households of different income levels in neighborhoods. Yet we know surprisingly little about the degree of economic mixing or segregation within US neighborhoods, certainly much less than we know about racial segregation.* [27]

Loretta Lees concluded that '*Social mix policies rely on a common set of beliefs about the benefits of mixed communities, with little evidence to support them and a growing evidence base that contradicts the precepts embedded in social mix policies…*'[28] A similar conclusion was provided by Andersson et al, including the suggestion that efforts towards mixing are even counter-productive.[29] Urban Studies scholar George Galster argues that mixing communities on *principle* as a policy initiative is '*too crude a mantra*' given the difficulty in weighing benefits and disadvantages for those involved.[30]

Perhaps no one has accused mixed-community policy of being belief or principle-based in nature more explicitly than Paul Cheshire of the *London School of Economics*. In his scholarly address on the subject, Cheshire refers to the practice of mixing as a '*faith-based displacement activity*'.[31] In his introduction, Cheshire writes:

> *We behave and apply policies as if it were a fact that the separation of different types of people and households into distinct and segregated neighbourhoods generated specific social costs, additional to those generated by inequality itself. But careful examination of the evidence suggests that such policies are more a matter of faith than anything else.*[32]

Consonant with Cheshire's argument, there exist several underlying reasons for labeling the guiding principle of working towards social integration as 'faith based'. Social integration efforts have been known to have a different or even opposite effect.[33] In addition to defining the proper 'mix' that should comprise a particular neighborhood or community[34], another problem with social mixing efforts is that neighborhood effects lead to differing results that can equally point towards support or suspicion, depending upon one's criteria. Furthermore, because de-concentration of poverty is such a popular method of creating social integration, dispersal methods often leave the existing neighborhood in worse condition than before, due to the methodological flaw[35] of (re)moving the more advantageous and resourceful members from the community, i.e., the 'cream of the crop'. Dispersal efforts, where low-income families are placed within middle- to higher-income neighborhoods, have shown outcomes that reveal the dispersed families are less likely to engage the social resources around them thus resulting in lower social capital.[36] Another important consideration to add to the tenuous nature of mixed income integration as it relates to housing is provided by John Calmore. Calmore, an African American and former law professor, called into question mainstream efforts towards social, and racial, integration. Citing a lack of integrity in this approach, he criticized '*spatial equality*' as a euphemism for '*white ethnocentrism*'.[37] Calmore provides the sobering reminder that efforts to socially integrate often come at the expense of cultural and ethnic identity, thus complicating the already muddled subject of *how* to socially integrate communities.

Thus, a collection of various arguments can be presented both for and against the plausibility of organized social integration. Whether it is Cheshire's clear outlining of the expensive nature of economically unsubstantiated integration policy such as MTO or John Calmore's disdain for racial condescension in 'spatial equality', the differing arguments are often packaged in economic and utilitarian language where teleological ends of individual utility and cost effectiveness dictate the direction taken on these important issues. This is not to suggest that such language is invalid, indeed, it isn't. Research informed policy feeds off such measures daily and is often the *raison d'etre* for the volumes of legislation that seek to shape, mold, and organize our contemporary society.

ISBN 978-0-9566139-1-2

The Ethics of Mixed Communities: A Theological Exploration

Since 1948, the BBC has hosted a series of radio lectures entitled the *'Reith Lectures'* that deal with '*significant contemporary issues, delivered by leading figures from relevant fields*'.[38] In 2009, the guest lecturer was Michael Sandel, Professor of Government at Harvard University. Among other lectures, Sandel offered thoughts on 'markets and morals', with his primary theme being that 'markets leave their mark'. In other words, markets tend to change the nature of the people, places and things they encounter.[39] Sandel echoed this sentiment in his 2009 book *Justice* where he points out that creating markets for otherwise aesthetically valued goods such as pregnancy/surrogacy, reading a book, civic virtue, etc., reduces the value of such goods to mere utility or money.[40]

Although this idea is not new[41], it was intuitive of Sandel to address this topic and in doing so validate it as a 'contemporary' issue. In his book, he concludes his chapter on markets with the question: '*are there certain virtues and higher goods that markets do not honor and money cannot buy?*'[42] Here, we might extend this idea with a related question of our own: is social integration—the idea of life lived together—a 'higher good' that cannot be reduced to an equation in gross utility? This is an important question, and one that the Christian faith tradition is uniquely positioned to answer. This last statement is based on two primary premises of traditional Christianity. The first is a distinction between value and optimality.[43] Christianity makes a distinct break from liberal traditions of the 'right' preceding the 'good'.[44] Second, the Christian faith tradition offers an overt trajectory of egalitarian relationships where socio-economic, racial, and cultural barriers are removed. While the faith tradition is not limited to these two distinctions, they provide a unique lens by which to re-appraise the 'higher goods' of social integration, thus extending the operative criteria we utilize to judge the issue of segregation, a phenomenon that has come to symbolize American social society.

In September of 2001, TIME magazine declared Duke Theologian Stanley Hauerwas 'American's best theologian'. Hauerwas responded by saying that 'best' was not a theological category; faithfulness was.[45] Here, Hauerwas captures an important distinction of the Christian faith, as 'best' is most often used as an economic term. The term implies efficiency, maximization, convenience, and effectiveness. However, the bible is clear that faithfulness cannot cling to such terms in the lexicon and practice of a disciple. Jesus' idea of faithfulness involves counter-intuitive propositions such as the '*last shall be first*'[46], becoming like '*little children*'[47], and the idea that the great will be servants and slaves to others.[48] Indeed, he showers blessings on those who are poor, those who mourn, the meek, the merciful, and the peacemakers; characteristics we typically don't celebrate in our modern western culture.[49] Such directives not only defy our intuition to do what is most economic (i.e., what is *best*), but are often found to be inefficient, inconvenient, and ineffective relative to our normal application of these terms.[50] To summarize, the behavior associated with traditional Christianity is not prized for its efficacy.

Thus, a commitment to the teaching of Jesus, a commitment to being faithful, is an exercise in reorienting mind and spirit to an alternative narrative. This gives the Christian faith tradition a natural link to the discipline of ethics. Ethics, often understood as the task of arriving at what is right or moral, is distinctly concerned with our social interactions. As Richard Longenecker writes, '*It is necessary to remind ourselves that ethics in the Bible are always set in relational contexts.*'[51] Indeed, when Jesus was asked to provide the greatest of the laws, he responded:

> *'You shall love the Lord your God with all your heart, and with all your soul, and with all your mind.' This is the greatest and first commandment. And a second is like it: 'You shall love your neighbor as yourself.'*[52]

It would be wrong, however, to construct the Bible as some kind of ethical handbook.[53] Rather, a biblical approach to ethics would be to hold ethical systems in tension. As mentioned, utilitarian outcomes of monetary efficiency and the maximization of gross happiness have been the primary bolts upon which the door of social integration swings open or closed. However, utilitarian ethical judgments claim that society is rightly ordered when the greatest net balance of satisfaction is summed over all the individuals.[54] However, this fails to recognize the inequality amongst individuals and even communities. For example, the classical idea of utilitarian economics is well summarized in the concept of 'pareto optimality', where a situation is optimal if at least one member can gain without the least

ISBN 978-0-9566139-1-2

member being made worse off. Such a term masquerades as being just and echoes the orthodox Millsian idea of what it means to be 'free'.[55] However, this principle can come '*hot from hell*', as stated by political philosopher Amartya Sen, because '*a state can be pareto optimal with some people in extreme misery and others in misery so long as the miserable cannot be made better off without cutting into the luxury of the rich.*'[56] Describing pareto optimality, Economist Donald Hay writes: '*Clearly if all such improvements are carried through, a point will be reached at which no citizen can be better off except at the expense of others.*'[57] Further, this formula for validating what is and is not successful in the market is based on a specific outcome while displaying a fundamental indifference to ethics insofar as being concerned with what is right in and of itself.[58]

The Christian faith tradition, and its commitment to faithfulness, divorces itself from this narrow methodology of determining social arrangements. When utilitarian summations of overall happiness are supplemented with the ideals of faithfulness towards God and towards neighbor, this provides reason for pause on several levels. First, the Judeo-Christian ethic of taking up the cause of the weak and marginalized must be considered. Justice, as stated in Isaiah 9:7, was a call to defend the weak and to seek solutions to violence and oppression.[59] Second, a Christian ethic aims towards responsible stewardship that is both glorifying to God and advantageous to others. Such distribution of resources should be accompanied by 'eagerness' and identified by 'fair balance'. The apostle Paul in his letter to the church in Corinth writes:

> *I do not mean that there should be relief for others and pressure on you, but it is a question of a fair balance between your present abundance and their need, so that their abundance may be for your need, in order that there may be a fair balance. As it is written, 'The one who had much did not have too much, and the one who had little did not have too little.'*[60]

Taking up the cause of the weak and demonstrating generosity of resources and power under our stewardship are distinctly Christian features that represent faithfulness to God and to neighbor. To act on these imperatives is often not an exercise in maximizing gross utility. However, incorporating these ideals into our thought exercises on social integration provides a more comprehensive perspective to our deliberation on what it 'best' as it relates to our social arrangements.

The second premise of the Christian faith tradition by which to examine the greater question of virtue and higher good as it relates to social integration is the Christian trajectory of egalitarian relationships where various segregating barriers are removed. As discussed, the Christian tradition, in its DNA, makes a break from the liberal tradition of the right preceding the good. Rather, more consonant with the Christian faith tradition would be the submission that determinations of justice are dependent on prior conceptions of what is good for humanity.[61] This, biblically speaking, is achieved in two separate but relevant declarations. The first is God's directive that life is not to be lived out in solitude: '*Then the Lord God said, "It is not good that the man should be alone."*'[62] Second, there is a clear line of ideology in the new covenant where community and solidarity, regardless of ethnic, cultural, or socio-economic status, are conceived as being the standard for deliberations in social ethics. For our purposes, my focus will be on the latter declaration.

To say that community and solidarity were distinctions of the early church is, in many ways, an understatement. Such an idea is difficult to comprehend in today's culture where rights-based liberalism perpetuates the ever-increasing individuality of our society and threatens the social fabric of our communities.[63] David Atkinson, acknowledging that justice, inclusion, and interdependence are at the heart of the biblical view of community, references Paul's text in Galatians 3:28 as an ethical summary of the life of the Christian church.[64] It reads: '*There is no longer Jew or Greek, there is no longer slave or free, there is no longer male and female; for all of you are one in Christ Jesus.*'[65] Richard Longenecker, who refers to this text as '*the most forthright statement on social ethics in the New Testament*', points out that this verse was more than an '*idiosyncratic notion of Paul*', but rather, '*should be taken as a confession of first-century Christians more generally—a confession included within the baptismal liturgy of the early Church which proclaims both a new status in Christ before God spiritually and new relationships between believers socially.*'[66]

ISBN 978-0-9566139-1-2

This idea is prevalent throughout the New Testament, where varying social barriers are replaced with the inclusive ethic of Jesus who breaks down '*the dividing wall of hostility*'.[67] This opens the doorway to a new community with distinct characteristics. Such characteristics include a selfless community through the emulation of Jesus. As New Testament scholar Richard Hayes remarks: '*The fundamental norm of Pauline ethics is the christomorphic life. To imitate Christ is also to follow the apostolic example of surrendering one's own prerogatives and interests.*'[68] Christian community stresses 'being' over 'having' where the community ethic is not contingent on performance and prestige.[69] This community is inclusive: women were intricately involved and played a decisive role in the ministry of Jesus against the backdrop of an otherwise patriarchal culture in a time where they were considered '*too emotional or too illiterate*'.[70] Slaves were no longer to be viewed as servants, but were now considered '*beloved brothers*'.[71] Cultural identity was analogous to social identity and highly associated with social outcomes. Yet the 'new self', identified in the community of faith, was transformed to where '*there is no longer Greek and Jew, circumcised and uncircumcised, barbarian, Scythian, slave and free; but Christ is all in all.*'[72] Finally, socio-economic barriers were removed as it related to first century integration. Indeed, the followers of Jesus ranged from the affluent to the poor and uneducated.[73] Believers found it necessary to redistribute possessions among themselves so that none had need. Counterintuitive to our modern affair with meritocratic ideology, early followers of Jesus sold their possessions and distributed the proceeds among themselves.[74] Acts 4:34 reads: '*There was not a needy person among them, for as many as owned lands or houses sold them and brought the proceeds of what was sold.*'

The community ethic of the Christian faith tradition reorients our perceptions towards living apart. Beyond utility and economic efficacy, a theological appeal to social integration is based on the moral ethic of transforming existing social roles so that there is equality among believers. Equally loved and equally valued, this new community found their identity in the person of Jesus and in their relationships with each other—relationships characterized not by culture, ethnicity, wealth or status, but rather the idea that 'all are one in Christ'.

CONCLUSION: The Future of Social Integration

Thus far, I have provided reflection on the idea of social integration and how a theology of integration might serve to inform and enhance our otherwise utilitarian perspective relative to this issue. At this point, it is valuable to provide deliberation for the implications of such reflection on the greater discussion of spatial equality and social integration. The purpose of this deliberation is not to suggest a singular prescriptive answer to this otherwise difficult issue. Rather, as mentioned earlier, my hope is to expand the ethical boundaries of the social integration discussion.

One pertinent area of discussion that arises out of reflection on theology and mixed communities is the idea of 'contact theory'. This, simply put, asserts that having contact with people from other groups can reduce prejudice.[75] This introduces the idea that if whites and blacks were less racially isolated, they might assess race problems differently and work together towards broader-based solutions.[76]

If contact is a necessary tool for scaling the dividing wall of segregation, the Christian faith tradition has both the epistemological starting point for contact and the early church tradition of its practice. I mention the former because Christian's believe that God sent his son, Jesus, to the world to redeem mankind.[77] The latter aforementioned element naturally flows from the acceptance and emulation of Jesus' example of proximity. In Luke 10, the famous parable of the 'good Samaritan' provides insight into the idea of boundary and neighborliness in terms of Christian ethics and morality. Jesus, being questioned by a lawyer, is asked '*And who is my neighbor?*' Jesus addresses the question by offering the famous parable where two men, a priest and a Levite, pass by a wounded and dying man '*on the other side of the road*'. However, a Samaritan '*came near him*' and provides him bandaging, shelter, and additional care. Jesus finishes his parable with the question: '*Which of these three, do you think, was a neighbor to the man who fell into the hands of the robbers?*'[78] Here Jesus avoids the descriptive question of 'who is my neighbor' and rather provides a prescriptive 'here is how to be a neighbor'.

ISBN 978-0-9566139-1-2

Thus, if contact theory is a part of the deliberation encompassing the discussion of mixed communities, then proximity, as a natural element of contact, must be addressed. The Christian faith tradition exposes the erroneous and circular process of constructing housing policy within the vacuum of a fixed neighborhood where exposure to different races, traditions, culture and socio-economic status is limited and the winds of housing policy blow towards maintaining this segregated establishment. If we accept the propositions of contact theory, can we truly base integration policy on utility and efficacy when we are already grossly segregated? Should we not aim towards advancements of integration and then make such assessments, as segregation can only serve to substantiate and thus exacerbate its own existence? The ethic of the Christian faith tradition beckons us to suspend our limited criteria of utility based outcomes, calling for overtures towards greater proximity and thus contact amongst otherwise segregated communities based upon the moral impetus that there is something humane, ethical, spiritual, about life lived together.

I have presented an appeal to expand the repertoire of evaluative criteria as it relates to the discussion and deliberation of mixed communities. Further, I have provided the framework for a theological ethic by which to assess this issue and its value for our contemporary society. One might rightly speak to the spiritual or religious nature of such an ethic and question its overall place and value in the trenches of public discourse. However, I would assert that social integration is an integral part of a greater discussion of justice, and questions regarding societal justice are at the heart of the Christian faith tradition. Doug Ottati, speaking to this issue, writes:

> *Why discuss justice in an explicitly theological frame of reference? Well, if it proves illuminating, why not? It seems self-defeating in this remarkably religious age to exclude theological conversation partners. In addition, we may note that Augustine, Thomas Aquinas, Francisco de Vitoria, and Francisco Suarez—theologian all—did much to advance discussion about justice and war in the West. Walter Rauschenbusch, Reinhold Niebuhr, and Martin Luther King, Jr.—all theologians once again—made signal contributions to discussions of what justice might require in and of 20th-century America.* [79]

There is value in assessing the moral implications of contemporary housing policy in addition to issues of utility and economic efficacy. I submit that this, in concert with the recent sentiments of *The Economist*, is a more appropriate measure for progress and prosperity. To exclude moral ideals and appeals in the discussion of spatial equality and social integration would be to betray the nature of community, which finds its meaning embodied in relationship and support—'higher goods' that cannot be limited to an economic understanding.

REFERENCE

[1] Massey, Douglas S., and Nancy A. Denton. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, Massachusetts: Harvard UP, 1993. Print. (Page 20)
[2] Ibid., page 20.
[3] Ibid., page 21.
[4] Ibid., pp. 3-4.
[5] Massey, Douglas S., and Nancy A. Denton. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, Massachusetts: Harvard UP, 1993. Print. (Page 67)
[6] Calmore, John O. "Spatial Equality and the Kerner Commission Report: A Back-to-the-Future Essay." *North Carolina Law Review* 71 (1993): 1487-518. Print. This is an astonishing statistic given the fact that there were approximately 30 million African Americans in the US as of the 1990 Census (See http://factfinder.census.gov/servlet/QTTable?_bm=y&-qr_name=DEC_1990_STF1_QTP1D&-geo_id=01000US&-ds_name=DEC_1990_STF1_&-_lang=en&-format=&-CONTEXT=qt). Similar to the 1990s, the 2000 census continues to find African Americans segregated in metropolitan areas (See http://www.censusscope.org/us/map_nhblack.html).
[7] Logan, John, and Lewis Mumford Center. *Ethnic Diversity Grows, Neighborhood Integration Lags Behind*. Rep. 2001. Print. (page 1)
[8] Massey, Douglas S., and Nancy A. Denton. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, Massachusetts: Harvard UP, 1993. Print. (Page 77)

ISBN 978-0-9566139-1-2

[9] The best example of this would be Fair Housing Act signed in 1985 which included the facilitation of forcing communities to provide a "fair share" of affordable housing for low to moderate income families—families that were otherwise prohibited from the community through exclusionary zoning tactics.
[10] Logan, John, and Lewis Mumford Center. *Ethnic Diversity Grows, Neighborhood Integration Lags Behind.* Rep. 2001. Print. (page 1)
[11] Blasius, Jorg, Jurgen Friedrichs, and George Galster. "Introduction: Frontiers of Quantifying Neighbourhood Effects." *Housing Studies* 22.5 (2007): 627-36. Print. (Page 627)
[12] Such a ruling, based more on moral ideals than constitutional precedence or economic dependency, was one of the more extraordinary aspects of the Mount Laurel exclusionary zoning case. For comprehensive assessments of the Mt. Laurel case, see Kirp, David L., John P. Dwyer, and Larry A. Rosenthal. *Our Town: Race, Housing, and the Soul of Suburbia.* New Brunswick, New Jersey, and London: Rutgers UP, 1997. Print. & Haar, Charles M. *Suburbs Under Siege: Race, Space, and Audacious Judges.* Princeton, NJ: Princeton UP, 1996. Print.
[13] Kirp, David L., John P. Dwyer, and Larry A. Rosenthal. *Our Town: Race, Housing, and the Soul of Suburbia.* New Brunswick, New Jersey, and London: Rutgers UP, 1997. Print. (Page 94)
[14] On the About Us section of their website, The Economist cites a commitment to "free trade and free markets." See "About Us." The Economist. Web. 19 May 2010. <http://www.economist.com/help/DisplayHelp.cfm?folder=663377#About_The_Economist.cfm?folder=663377#About_The_Economist>.
[15] This defense has taken on several forms, most notably from market defenders who insist that the financial crisis is just as much of a regulatory failure as it is a market failure.
[16] "Onwards and Upwards." *The Economist* 19th Dec. 2009: 37-40. Print.
[17] Ibid., page 40.
[18] Ibid.
[19] Kearns, Ade, and Phil Mason. "Mixed Tenure Communities and Neighbourhood Quality." *Housing Studies* 22.5 (2007): 661-91. Print. (Page 664)
[20] Massey, Douglas S., and Nancy A. Denton. *American Apartheid: Segregation and the Making of the Underclass.* Cambridge, Massachusetts: Harvard UP, 1993. Print. (Page 85)
[21] Kirp, David L., John P. Dwyer, and Larry A. Rosenthal. *Our Town: Race, Housing, and the Soul of Suburbia.* New Brunswick, New Jersey, and London: Rutgers UP, 1997. Print. (Page 9)
[22] Ibid., page 88.
[23] Ibid., page 9.
[24] Flippen, Chenoa. "Unequal Returns to Housing Investments? A Study of Real Housing Appreciation among Black, White, and Hispanic Households." *Social Forces* 82.4 (2004): 1523-551. Print. (Page 1524)
[25] Vartanian et al (2007) have provided a comprehensive account of neighborhood disadvantage for segregated African Americans. See Vartanian, Thomas P., Page W. Buck, and Philip Gleason. "Intergenerational Neighborhood-Type Mobility: Examining Differences Between Blacks and Whites." *Housing Studies* 22.5 (2007): 833-56. Print.
[26] Kearns, Ade, and Phil Mason. "Mixed Tenure Communities and Neighbourhood Quality." *Housing Studies* 22.5 (2007): 661-91. Print. (Page 664)
[27] Hardman, Anna, and Yannis M. Ioannides. "Neighbors' Income Distribution: Economic Segregation and Mixing in US Urban Neighborhoods." *Journal of Housing Economics* 13 (2003): 368-82. Print. (Page 369)
[28] Lees, Loretta. "Gentrification and Social Mixing: Towards an Inclusive Urban Renaissance?" *Urban Studies* 45.12 (2008): 2449-470. *Http://usj.sagepub.com.* Glasgow University. Web. 11 May 2009. <http://usj.sagepub.com/cgi/content/abstract/45/12/2449>. (Page 2463)
[29] Andersson, Roger, Sako Musterd, George Galster, and Timo M. Kauppinen. "What Mix Matters? Exploring the Relationships between Individuals' Incomes and Different Measures of Their Neighbourhood Context." *Housing Studies* 22.5 (2007): 637-60. Print.
[30] George Galster as quoted by Kearns, Ade, and Phil Mason. "Mixed Tenure Communities and Neighbourhood Quality." *Housing Studies* 22.5 (2007): 661-91. Print. (Page 688)
[31] Indeed, this was the title of his paper presented at the 54th Annual North American Meetings of the Regional Science Association International in 2007: Cheshire, Paul. *Mixing Communities: A Faith-Based Displacement Activity?* Proc. of 54th Annual North American Meetings of the Regional Science Association International: Savannah 8-10 November 2007. London: London School of Economics, 2007. 1-33. Print.
[32] Ibid., 3.
[33] An example of gentrification producing an unintended consequence is cited taking place in the Netherlands. Lees writes: "Uitermakr et al. (2007) are clear that an influx of middle-class residents into a disadvantaged neighbourhood does not increase social cohesion, rather the contacts between low-income and higher-income households tend to be superficial at best and downright hostile at worst" (Lees, 2008, p. 2456). Rowland Atkinson cites a similar study where gentrification displaced low-income homeowners from their area. See CNR (Center for Neighbourhood Research) review of the impacts of gentrification, Atkinson, (2002).

ISBN 978-0-9566139-1-2

[34] Andersson et al (2007) write: "However, such [programs for integration] seldom make clear what mix is desirable and appropriate, only that mix is good. Therefore, clarifying what mix matters is seen her as an important task for social science research" (Andersson, 2007, p. 656).
[35] This is Atkinson's term. It is a "flaw" in the sense that the selection process for those worthy of vouchers will choose, and thus displace, those "that were already likely to do better than the wider group of residents in pre-existing areas of black and poor neighbourhoods" (Atkinson, 2005, p.20).
[36] Ibid., page 22 (Atkinson)
[37] Calmore, John O. "Spatial Equality and the Kerner Commission Report: A Back-to-the-Future Essay." *North Carolina Law Review* 71 (1993): 1487-518. Print. (pp. 1505-1506)
[38] "BBC - BBC Radio 4 Programmes - The Reith Lectures." *BBC - Homepage*. Web. 25 May 2010. <http://www.bbc.co.uk/programmes/b00729d9>.
[39] Ibid., for specific link to speech, see http://www.bbc.co.uk/programmes/b00kt7sh for the lecture titled "Markets and Morals."
[40] Sandel, Michael J. *Justice: What's the Right Thing to Do?* New York: Farrar, Straus and Giroux, 2009. Print. (Page 98)
[41] This idea carries all the way back to the time of Aristotle and his exercise of determining the *telos* of a particular social practice, e.g., its purpose, end, or good. Sandel describes this as Aristotle working to "identify the norms appropriate to social practices by trying to grasp the characteristic end." (Page 98)
[42] Ibid., Page 102.
[43] One might respond that this tradition appears no different than Kant's deontological project, where something is "right" regardless of its usefulness as long as it can be universally accepted (i.e., the *Categorical Imperative*). However, the Christian tradition distinguishes itself by claiming a covenant ethic between man and man and man and God. Robin Lovin, in contrasting this tradition with the natural law tradition, writes: "Natural law locates the moral rules in an order that is part of reality. Covenant thinking locates them in a relationship of faithfulness between people or between individuals and God." Further, Lovin acknowledges that a covenant relationship often requires more specific things of us and can impose obligations that are not considered universal. See Lovin, Robin W. *Christian Ethics: An Essential Guide*. Nashville: Abingdon, 2000. Print. (Page 50)
[44] The term "right" in this sense is understood in its traditional liberal context, i.e., the "rights" of each human being. This is not to be confused with "right" as an ethical judgment, i.e., we should strive to do what is "right."
[45] McCarthy, Coleman. "I'm a pacifist because I'm a violent son of a bitch." *A profile of Stanley Hauerwas*. (The Progressive, April, 2003).
[46] Matthew 19:30
[47] Matthew 18:3
[48] Matthew 20: 26-28
[49] Matthew 5:3-11
[50] While in language the scriptures and the Christian faith tradition in general make a distinction between 'best' and 'faithful', this practice is not universally held by the church. Biblical manipulations creating "prosperity gospel" and other perversions of the faith tradition take a clear line of departure from this aforementioned thought.
[51] Longenecker, Richard N. *New Testament Social Ethics for Today*. Grand Rapids, Mich.: W.B. Eerdmans, 1984. Print. (pp. 13-14)
[52] Matthew 22:37-39
[53] Gorringe, Timothy J. *Capital and the Kingdom: Theological Ethics and Economic Order*. Maryknoll, New York: Orbis, 1994. Print. (Page 19)
[54] Rawls, John. *A Theory of Justice*. New Delhi: Universal Law, 1971. Print. (Page 22)
[55] Mill defines freedom in this way: "The only freedom which deserves the name is that of pursuing our own good in our own way, so long as we do not attempt to deprive others of theirs, or impede their efforts to attain it." As quoted by Bauckham, Richard. *God and the Crisis of Freedom: Biblical and Contemporary Perspectives*. Louisville, KY: Westminster John Knox, 2002. Print. (Page 27)
[56] Amartya Sen as quoted by Gorringe, Timothy J. *Capital and the Kingdom: Theological Ethics and Economic Order*. Maryknoll, New York: Orbis, 1994. Print. (Page 35)
[57] Hay, Donald A. *Economics Today: A Christian Critique*. Vancouver: Regent College, 1989. Print. (Page 154)
[58] Cramp, Anthony B. "What Kind of Financial System Do We Want?" *Ethics and Finance Group Paper No.2* (1995): 1-14. Print. (Page 10)
[59] Gorringe, Timothy J. *Capital and the Kingdom: Theological Ethics and Economic Order*. Maryknoll, New York: Orbis, 1994. Print. (Page 14)
[60] II Corinthians 8:13-15
[61] Bell, Daniel Deliberating: Justice and Liberation. Article found in Hauerwas, Stanley, and Samuel Wells. *The Blackwell Companion to Christian Ethics*. Malden: Blackwell, 2006. Print. (pp. 182-195)
[62] Genesis 2:18

ISBN 978-0-9566139-1-2

[63] This is the argument of Harvard professor Robert Putnam. In his book *Bowling Alone*, he traces a decline in "social capital" over the last few decades and argues that Americans need to reconnect to one another, less they threaten the value of their social fabric and the overall nature of community. See Putnam, Robert D. *Bowling Alone: the Collapse and Revival of American Community*. New York: Simon & Schuster, 2000. Print.
[64] Atkinson, David. "Community Ethics." *New Dictionary of Christian Ethics and Pastoral Theology*. Leicester: Inter-Varsity, 1995. 108-115. Print. (Page 113)
[65] Galatians 3:28
[66] Longenecker, Richard N. *New Testament Social Ethics for Today*. Grand Rapids, Mich.: W.B. Eerdmans, 1984. Print. (Page 33)
[67] Ephesians 2:14
[68] Hayes, Richard B. *The Moral Vision of the New Testament: A Contemporary Introduction to New Testament Ethics*. New York: Harper Collins, 1996. Print. (Page 46)
[69] Atkinson, David. "Community Ethics." *New Dictionary of Christian Ethics and Pastoral Theology*. Leicester: Inter-Varsity, 1995. 108-115. Print. (Page 109)
[70] Witherington, Ben. *What Have They Done with Jesus? Beyond Strange Theories and Bad History--Why We Can Trust the Bible*. San Francisco: Harper Collins, 2006. Print. (Page 18)
[71] Philemon verse 16
[72] Colossians 3:11
[73]Affluent followers included Joanna, the wife of Herod's steward Chuza, Matthew the Roman tax collector, and Luke the physician. Other individuals that ran the stereotypical spectrum as followers of "the way" were uneducated fisherman, prostitutes, centurions, and other marginalized members of mainstream society.
[74] Acts 2:45
[75] Ibid., Page 106.
[76] Ibid., Page 132.
[77] Jesus was referred to as "Emmanual" meaning "God with us."
[78] Luke 10: 25-37
[79] Ottati, Douglas F. "What Reformed Theology in a Calvinist Key Brings to Conversations about Justice." *Political Theology* 10.3 (2009): 447-69. Print. (pp. 1-2)

*NOTE: All scripture referenced is from the New Revised Standard Version (NRSV)

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

ENGAGING WITH DIVERSE APPROACHES TO HUMAN SKELETAL ADAPTATION: CLIMATE, ACTIVITY, AND INTEGRATION

Thomas DAVIES

PhD Student, Leverhulme Centre for Human Evolutionary Studies, University of Cambridge, United Kingdom
td274@cam.ac.uk

Keywords: Human Adaptation, Morphology, Biomechanics, Climate, Integration

ABSTRACT

The human skeleton is a complex biological system known to adapt to its environmental context. Two dominant paradigms have emerged in the study of human skeletal adaptation. Firstly, that human skeletons are adapted to the climates in which they mature, most notably seen in limb proportion and body size trends with latitude according to Bergmann and Allen's ecogeographic rules. Secondly, that the human skeleton is plastic to the mechanical forces to which it is exposed during an individual's life. A body of research over the past 20 years has exposed plastic components to skeletal traits including: long bone robusticity, curvature, and torsion. Research that engages with both of these paradigms, 'drawing in the margins', is rare.

The focus of research is often constrained to examining single traits from single perspectives. Whilst this is critical, new insights may be gleaned by exploring these boundaries drawn between climate and activity driven approaches, and in the relationships between traits within the skeleton. Two questions are addressed which seek to develop research into these areas: 1. Can a high resolution, 3-dimensional approach to skeletal adaptation reveal insights into the role of climatic effects upon limb bone morphology? 2. To what degree are traits both across the skeleton and within the limbs integrated in their responses to environmental and mechanical stimuli? The first question explores an area of crossover between the dominant paradigms that has recently emerged: impacts of body size parameters upon limb bone diaphyseal morphology. The second investigates further relationships between skeletal traits with a significant bearing on understanding climatic and behavioural influences upon the skeleton.

It is often at the juncture of different disciplines that new insights emerge, but even within a single, small discipline points of juncture must also be given due attention.

INTRODUCTION

Human Adaptation

Adaptation encompasses the ability of an organism to fit to, and succeed in, its environment. This can be either genetic adaptation by natural selection over many generations or the physiological, technological or behavioural adaptations of an organism to meet the specific demands of environmental stresses experienced during life.

Much attention given to explaining the human species has been focused on understanding the varying influences that genes and environments exert on the human phenotype (i.e. the observable characteristics or traits of an organism); traditionally termed as the nature versus nurture debate.[1] Regarding human skeletal morphology such debates can be traced back at least as far as Franz Boas in the early 1900s and his research showing that

ISBN 978-0-9566139-1-2

descendents of immigrants to the United States differed in morphology compared to that of their foreign-born parents, revealing a key role for environmental influences on skeletal morphology.[2] One recent attempt to draw together our understanding of the distinct processes that have acted upon our species is the colonizing ape hypothesis, which argues that hominin adaptations to unstable Pleistocene environments and cycles of colonization and dispersal throughout human evolutionary history brought about a selection for generalized biology and favoured significant phenotypic plasticity to buffer the genome from environmental stress.[3] This also raises the importance of understanding modern human biology as a means to develop our understanding of human evolution and now extinct hominin species. Only by exploring patterns of variation on a population and species level (i.e. within modern humans) will it be possible to develop our ability to interpret the often fragmentary and sparse fossil remains. Whilst genetic advances such as the recent publication of Neanderthal genome[4] represent key technological developments, the complete picture cannot be constructed without a similarly developed understanding of phenotype (the product of both genetics and environment).

Whilst the study of human skeletal adaptation is itself interdisciplinary, often drawing upon elements from fields as disparate as osteology, biomechanics, growth, ecology, evolution, genetics, biology, archaeology and ethnography, it is possible to identify two key areas of focus in the study of human postcranial skeletal adaptation.

The first is adaptation to the natural environment in its strictest sense, the means by which climate and temperatures impose strains upon the skeletal system. Relationships between body size and climate have been firmly established since the formation of Bergmann and Allen's ecogeographic rules back in the 19thCentury.[5,6] Roberts in 1953 demonstrated that these apply to humans[7], and Katzmarzyk and Leonard confirmed these findings fifty years later (though the smaller correlation they identified hints also at the additional roles of nutrition and plasticity in determining body size and proportions).[8] Patterns in skeletal limb proportions, both in modern humans and late Pleistocene hominins, reflecting climatic adaptation have been explored[9,10,11], and the most recent studies continue to fill in the population-specific detail.[12]

The second takes a behavioural or activity-based approach - adaptation to the physical environment - to develop the means by which the skeletal system adapts to forces to which it is exposed during life. Such biomechanical approaches developed in 1970s[13] but have received much attention since three key papers in the early 1990s.[14,15,16] Research in this field has shown that limb bone shaft morphology reflects the intensity and repetitiveness of the forces imposed upon it during life.[17,18] One illustration is Stock & Pfeiffer's contrasts between Andaman Islanders, who have a predominantly marine lifestyle and thus use their arms heavily in locomotion, and display reduced leg strength but increased arm strength in comparison to terrestrial foragers from South Africa.[19] Other traits within the limbs that have revealed signatures of phenotypic plasticity to behaviours are humeral torsion (angle of humeral head)[20] and long bone curvature.[21,22]

These two areas of research into skeletal adaptation have remained relatively distinct from one another. One of few significant elements of cross-over is a need to correct for body size in measures of skeletal strength, but this is used in a manner that isolates behavioural signals, rather than investigating interactions.[23] By standardizing for body size, remaining variation in strength is argued to be the result of behavioural differences (i.e. a plastic response to the forces imposed upon the limb during growth and development).

Whilst most research has been constrained within one or other of these schools, there have been a number of attempts to bridge the division, and take a more encompassing approach to skeletal adaptation. These developments have opened new opportunities to explore the marginal areas that overlap between climatic and behavioural approaches. Furthermore, integrating insights from both approaches into analyses holds the potential to increase our understanding of the overall framework of processes that act upon the human skeleton.

Integrating climatic and behavioural approaches

Churchill demonstrated that the evolution of the upper arm involves both particulate (i.e. independent) and integrated (i.e. covariation across the skeleton) components.[24,25] This raises the possibility that the morphology of the limbs (for example of long bone strength) may to an extent be influenced by variation in other elements of the skeleton (such as body shape). Thus the assumption that correcting limb bone strength for body mass reveals only behavioural information from the skeleton, may not be completely justified.

ISBN 978-0-9566139-1-2

In relation to long bone morphology, Weaver noted differences in the proximal femur (i.e. the top of the femur) between modern humans and Neanderthals were a secondary consequence of differences in climate induced body proportions.[26] Subsequently, Stock investigated the effects of both climate and behaviours on long bone shaft morphology, finding that climatic effects appear to exert more proximally (closer to the hip) whilst behavioural signatures are revealed at midshaft locations.[27] These results are in line with the theory that the proximal femur has greater medio-lateral (the horizontal axis parallel to body breadth) forces passing through it[28] or that distal segments might be exposed to a greater selection for tissue economy.[29] These findings suggest that different forces exert pressures differentially at various locations down the length of the limb. Thus a picture is building of long bone shaft morphology as playing a critical role at the synthesis of both climatic and behavioural influences on the skeleton. There is a need to explore further how climatic adaptation in overall skeletal size and shape, may exert integrated influences on limb bone strength.

Research is also starting to move beyond the constraints of examining single factors from single perspectives; there is much to be gained from understanding the relationships between different skeletal traits both across different skeletal elements and within the limbs. Two examples of work comparing traits within the limbs are Rhodes and Knusel's study comparing patterns in humeral torsion to long bone strength in Medieval skeletal samples[30] and De Groote's work on diaphyseal curvature discussed in relation to long bone strength.[31] Work is ongoing to investigate these relationships, and begin to develop approaches that take a more encompassing approach to understanding skeletal adaptations.

This study introduces a new method of quantifying limb bone strength and reports preliminary results addressing issues that seek to integrate both climatic and behavioural perspectives towards human skeletal adaptation. Preliminary results are explored that offer insights into two key areas surrounding the issue of how limb bones reflect external influences on skeletal anatomy. Firstly, the degree to which 3D approaches can reveal insights into climatic influences upon limb bone shaft morphology. And secondly, the extent to which traits across the limbs are integrated in their responses to such stimuli.

METHODOLOGY

Materials

For this study human skeletal remains were assessed from the Duckworth Collection, University of Cambridge. The sample consists of Australian aboriginals (n=13), Inuit from Ungava Bay and Greenland (n=7), Mori-ori skeletons from the Chatham Islands (n=2), an Iron Age sample from Maiden Castle, UK (n=11), and Badari Egyptian early agricultural sample (n=7).

Methods

This study utilizes a novel method for quantifying limb bone strength. This method involves 3-dimensional laser-scanning of bones to create a virtual reconstruction of the bone's surface morphology, from which cross-sectional morphology is subsequently extracted. 3D laser scans were taken using NextEngine 3D desktop scanner, shell alignment and processing was carried out in ScanStudio and Rapidform 2006 software, and subsequent processing to extract cross-sections and total area (TA) measurements using customized software. Cross-sections are extracted perpendicular to bone shaft for every 1% of bone length between 20-80% of bone length (measured from distal end).

This method mimics the external periosteal moulding approach in which only the periosteal (outer surface) contour is captured, thus recreating 'solid' cross-sections (Fig. 1). Stock and Shaw demonstrated high correlations between solid and medullary cross-sections, supporting the argument that the endosteal contour exerts minimal impact on bone strength in non-pathological cases based on the theoretical prediction that the further away bone mass is situated from the section centroid, the greater its role in resisting bending and torsional loading.[32,33]

ISBN 978-0-9566139-1-2

Figure 1: Example of solid femoral cross-sections (from left to right 20% to 80% sampled every 10% of bone length from distal end).

Where applicable, measures of cross-sectional strength (TA) were standardized using body mass estimates to correct for the effects of overall size.[34] Body mass was estimated using femoral head diameter (using the average of the three equations for body mass from femoral head diameter provided by Ruff *et al.*)[35] and stature using femoral length with Feldesman & Fountain's generic regression equation.[36]

DISCUSSION

Drawing in the margins of human skeletal adaptation

This discussion reports preliminary results, utilizing novel methodology, to bring the focus back into human skeletal adaptation in a broad context. As has been shown the two dominant approaches in this field are diverse, but in many cases lack engagement, these results demonstrate possible avenues towards exploring the overlapping margins of these two approaches.

Diaphyseal strength plots and evidence for climatic influence

Utilizing 3D laser scanner technology, makes it possible to relatively quickly establish measures of cross-sectional strength down the length of the limb, whereas previous studies have been restricted to key locations on the shaft (subtrochanteric; midshaft). Figure 2 plots the results for total area (TA) of cross-sections between 80-20% of bone length for four recent human populations.

The most obvious difference between the samples is in the overall strength of the limbs. It is difficult to draw any conclusions as to why this may be the case, however, because effects of body size, behavioural differences, or possible climatic influences may all be contributing to overall strength. A more interesting result, however, is in the difference in gradient in the plots between 40% and 80%. The Inuit and Maiden Castle samples appear to display a pattern of gradually increasing strength between 40-50% and 75%, which contrasts the comparatively consistent, flat or unchanging, strength of sections in this region of the limb for both Australian and Egyptian samples. Between 75-80% strength increases in all populations, but more strongly in the Inuit and UK samples. These two trends are interesting because they appear to be consistent with the climatic differences between the populations. It is plausible that wider body breadth, as expected in the Inuit and UK samples that lived at higher latitudes according to Bergmann's rule (to increase mass relative to surface area and thus retain more body heat), might be causing an increased strength in the proximal femur, perhaps by increasing medio-lateral loads in the femur. Whilst this would be consistent with Stock's findings of climatic effects on the proximal femur[37], alternative explanations may also explain these results. For example, the trend is also consistent both with increasing body mass and overall strength of the limb (as both Inuit and UK samples are also the largest in body mass and strongest overall (in TA) of the four) and so it is also possible the pattern may be a functional consequence of increased strength. Behavioural differences between the samples have also not been controlled for.

ISBN 978-0-9566139-1-2

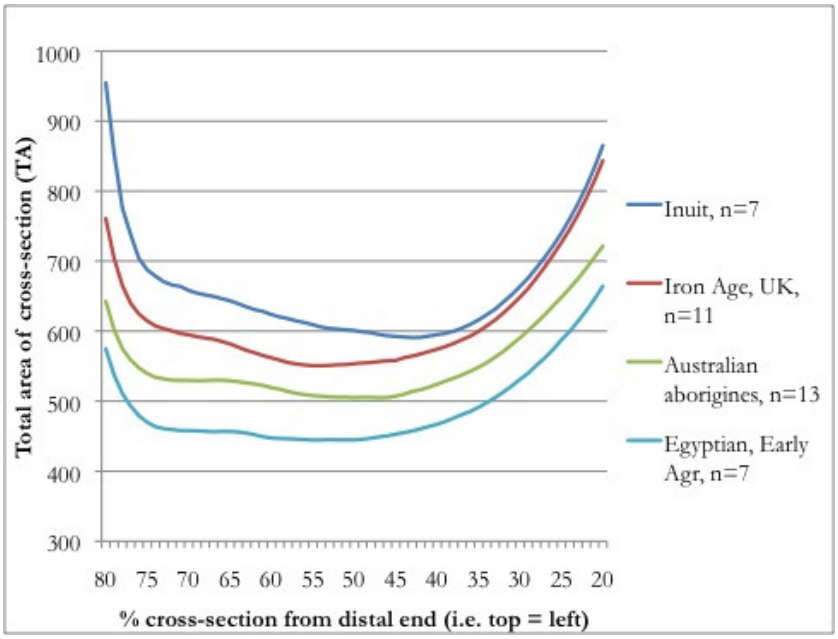

Figure 2: Average Total Area (TA) measurements between 80% and 20% of limb length for four recent human populations.

In order to show that these differences are not simply a function of body mass on the limb, Figure 3 plots the same data but with standardized TA values. It can be seen that the trends identified above relating to possible climatic adaptation (i.e. the strength gradient between 40 and 80%) can still be identified in the Inuit and UK samples, and thus that the trends are independent of body size. This is expected given that we are concerned here with difference in patterning of strength down the length of the limb and the standardizing process applied simply makes the same adjustment to all cross-sections for each individual. This suggests that body mass can be ruled out as an explanation, though it does not exclude the possibility that body mass has differential effects at different locations down the length of the limb, or that the pattern is caused by increased overall limb strength.

The biggest difference between the two plots (Figs. 2-3) is that by standardizing for body size the average strengths for the populations have been drawn closer together. The Inuit sample remains stronger, which is an interesting result as this sample might have been expected to be weaker in femur strength because of their marine locomotion. However, possible explanations of this result include smaller limb length or sex bias in the sample. Whilst it is difficult to draw firm behavioural comparisons at this stage in the analysis, the fact that the possible climatic trends identified persist after correction for body mass, suggests patterns of covariation across the skeleton merit further investigation (see below).

ISBN 978-0-9566139-1-2

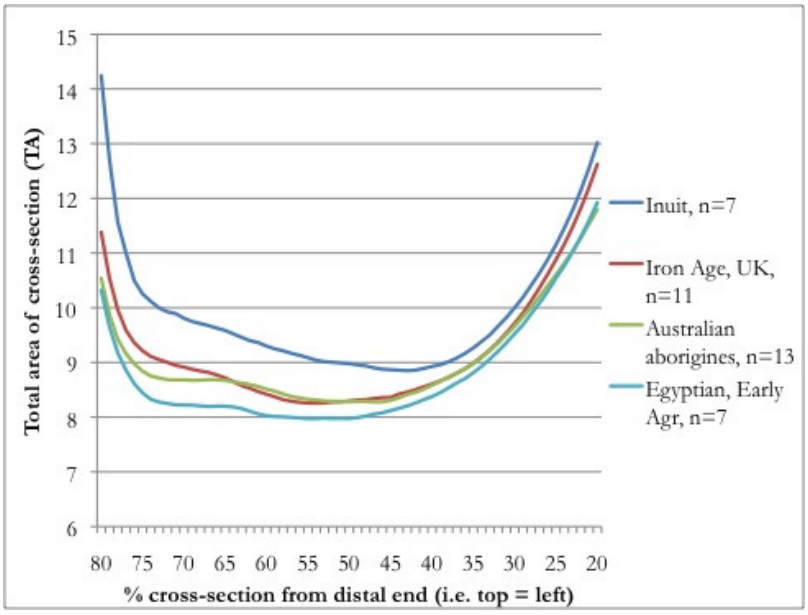

Figure 3: Average Standardized Total Area (TA) measurements between 80% and 20% of limb length for four recent human populations.

The approach taken here, utilizing 3D laser scanning to capture strength information down the length of the limb, dramatically increases the resolution of information attained from individual bones over traditional approaches. This has a number of advantages, notably in exposing variation across a single limb and in future application to interpreting hominin fragments where midshaft information may not be preserved.

Integration across the skeleton

The second question to be addressed is the degree to which skeletal traits are integrated in their responses to external stimuli. Here a detailed look is taken at cross-sectional strength (TA) against an estimate of relative body breadth, as a means of further increasing our understanding of the possible climatic trends identified in the previous section. Figure 4 plots the correlation between relative body breadth (bi-iliac breadth / stature) against total area (TA) for both the 50% and 80% cross-section levels. There is clearly a stronger correlation at the 80% level. This is a result that is consistent with Stock's conclusions of climatic effects upon the proximal femur.[38] This correlation represents a form a climatic adaptation (as body breadth is a key indicator of climatic adaptation[39]) on limb bone morphology. It may be expected that this increased strength represents a physiological plasticity to increased medio-lateral loads in individuals with wider bodies.

ISBN 978-0-9566139-1-2

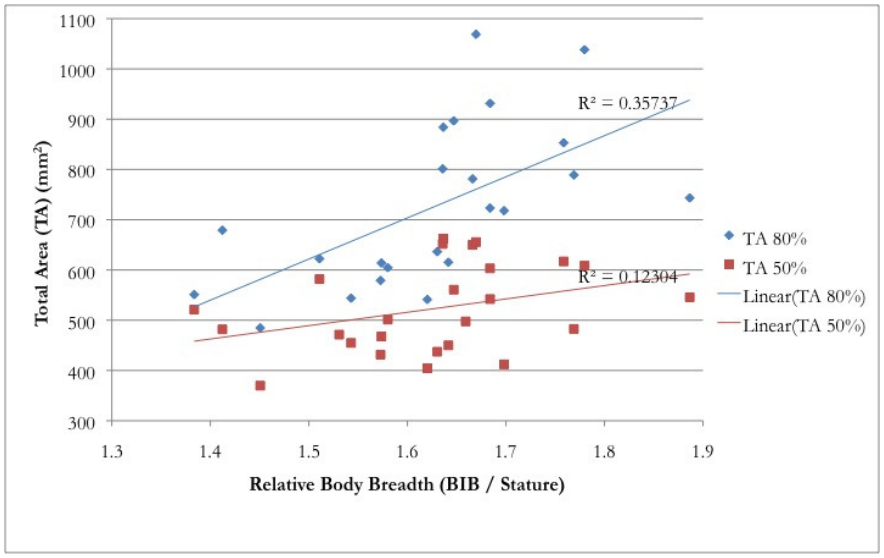

Figure 4: Correlations between relative body breadth and long bone strength (TA) for both proximal (80%) and midshaft (50%) cross-sections.

Examining this pattern further Figure 5 plots the R^2 values (percent of variance in TA explained by relative body breadth) for cross-section levels 80% through 20% (sampled every 1% of bone length). There is a decreasing trend in the correlation that persists from 80% down to at least the 65% level. This both illustrates that the correlation found for the 80% level is not an isolated result and implies that there may be a body breadth effect that continues to exert influence down a substantial portion of the femoral shaft. Supporting the climatic hypothesis for the trends in morphology identified in the previous section. These results represent preliminary findings and work is ongoing to increase the dataset and investigate these trends through further statistical analysis.

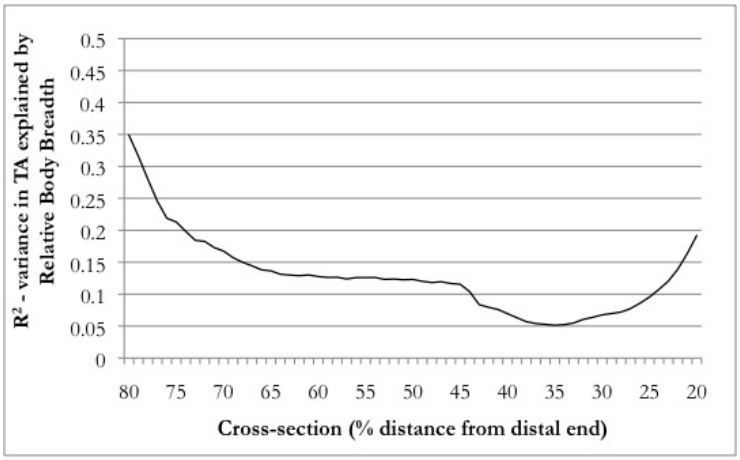

Figure 5: Plot of R^2 values showing the proportion of variance in TA explained by relative body breadth for each cross-section level 80-20% femur (n=25).

The results reported here investigate the correlations between relative body breadth and unstandardised TA values (i.e. raw cross-sectional area). Whilst therefore a body mass component will also play a significant role in the variation in these values, standardizing for body size does not alter the trends identified as standardization affects all cross-sections of each bone equally, and therefore has no effect in altering patterns in the relative

ISBN 978-0-9566139-1-2

strength of TA down the length of a bone (as demonstrated by Figs. 2-3 in previous section). The affect of standardizing for body size thus actually only acts to increase the extent of remaining variance explained by breadth for height, but does not alter the pattern across the limb.

It is unclear why the distal femur also demonstrates an increasing correlation with breadth for height, This indicates that at this early stage in the analysis one cannot be too confident in having identified a climatic effect and it must be noted that other factors such as the transition from cortical to trabecular bone within the femur, or behavioural contrasts between the populations, may have also acted to contribute to the apparent trend. Nevertheless, the proximal femur shows highest overall correlations, and a climatic explanation appears to be the most parsimonious at present. Body breadth for height appears to be increasingly correlated with diaphyseal strength nearer the top of the limb, further investigations will continue to isolate the degree to which this effect persists further down the bone. Whilst these results are preliminary, they illustrate the potential value in research that crosses the boundaries between behavioural and climatic approaches to skeletal adaptation.

CONCLUSION

In conclusion, taking a broad look at the range of studies of human skeletal adaptation offers one insight into the marked differences in approach that can develop between separate research areas, even within a comparatively small discipline. Whilst it is of course necessary and expected that detailed research be conducted within any single research area, these in turn provide scope for studies that develop within the marginal areas between them.

This paper has introduced a new approach that takes examines long bone strength across the length of long bones using 3D laser imaging. Whilst this approach is not without disadvantages (notably time consumed in scanning and data processing, and the restriction to capturing only the surface morphology), it offers the ability to extract rich data, and examine properties that are difficult to assess using traditional methods.

Preliminary results were reported demonstrating early avenues to exploring elements of the borderland between behavioural and climatic approaches. The results so far are consistent with previous findings but the methods applied here will ultimately allow for much greater level of information to be extracted. Work is ongoing to investigate deeper into these preliminary findings. This research is both critical in developing an integrated understanding of the human skeleton as an integrated biological system, and in furthering our ability to isolate particular signatures from skeletal material that can reveal glimpses into the lives and behaviours of past human populations, as well as extinct hominin species. Further research will be conducted both into these patterns, as well as into the ontogeny of these traits, which may provide the key to distinguishing the degree to which the processes impacting upon limb morphology are genetic or plastic.

It is often at the juncture of different disciplines that new insights emerge, but even within a single, small discipline points of juncture must also be given due attention.[40]

REFERENCES

[1] Ridley, M (2003). *Nature Via Nurture*. Fourth Estate.
[2] Boas, F. (1912). Changes in the Bodily Form of Descendants of Immigrants. *American Anthropologist*. 14(3): 530-562.
[3] Wells, J.C.K. & Stock, J.T. (2007). The Biology of the Colonizing Ape. *Yearbook of Physical Anthropology*. 50: 191-222.
[4] Green, R.E., Krause, J., Briggs, A.W. *et al.* (2010). A Draft Sequence of the Neandertal Genome. *Science*. 328: 710.
[5] Bergmann, C. (1847). Ueber die Verhaltnisse der Warmeokonomie der thiere zu ihrer grosse. Gottinger Studien 3: 595-708.
[6] Allen, J.A. (1877). The inflence of physical conditions in the genesis of species. Rad. Rev. 1: 108-140.
[7] Roberts, D.F. (1953). Body weight, race, and climate. *Am J Phys Anthropol*. 1: 533-558.
[8] Katzmarzyk, P.T. & Leonard, W.R. (1998). Climatic Influences on Human Body Size and Proportions: Ecological Adaptations and Secular Trends. *Am J Phys Anthropol*. 106: 483-503.
[9] Ruff, C.B. (1994). Morphological Adaptation to Climate in Modern and Fossil Hominids. *Yearbook of Physical Anthropology*. 37: 65-107.

ISBN 978-0-9566139-1-2

[10] Holliday, T.W. & Falsetti, A.B. (1995). Lower limb length of European early modern humans in relation to mobility and climate. *Journal of Human Evolution*. 29(2): 141-153.
[11] Holliday, T.W. (1997). Body proportions in Late Pleistocene Europe and modern human origins. *Journal of Human Evolution*. 32: 423-447.
[12] Holliday, T.W. & Hilton, C.E. (2010). Body Proportions of Circumpolar Peoples as Evidenced from Skeletal Data: Ipiutak and Tigara (Point Hope) Versus Kodiak Island Inuit. *Am J Phys Anthropol*.
[13] Lovejoy, C.O., Burstein, A.H., & Heiple, K.G. (1976). The biomechanical analysis of bone strength: a method and its application to platycnemia. *Am J Phys Anthropol*. 44: 489-506.
[14] Ruff, C.B., Trinkaus, E., Walker, A., & Larsen, C.S. (1993). Postcranial robusticity in Homo. I: Temporal trends and mechanical interpretation. *Am J Phys Anthropol*. 91(1): 21-53.
[15] Trinkaus, E., Churchill, S.E., Ruff, C.B. (1994). Postcranial robusticity in Homo. II: Humeral bilateral asymmetry and bone plasticity. *Am J Phys Anthropol*. 93(1): 1-34.
[16] Ruff, C.B., Walker, A., Trinkaus, E. (1994). Postcranial robusticity in Homo. III: Ontogeny. *Am J Phys Anthropol*. 93(1): 35-54.
[17] Stock, J.T. & Pfeiffer, S. (2004). Long bone robusticity and subsistence behaviour among Later Stone Age foragers of the forest and fynbos biomes of South Africa. *Journal of Archaeological Science*. 31: 999-1013.
[18] Shaw, C.N. & Stock, J.T. (2009). Intensity, Repetitiveness, and Directionality of Habitual Adolescent Mobility Patterns Influence the Tibial Diaphysis Morphology of Athletes. *Am J Phys Anthropol*.
[19] Stock, J.T. & Pfeiffer, S. (2001). Linking structural variability in long bone diaphyses to habitual behaviors: foragers from the southern African Later Stone Age and Andaman Islanders. *Am J Phys Anthropol*. 115(4): 337-48.
[20] Rhodes, J.A. (2006). Adaptations to Humeral Torsion in Medieval Britain. *Am J Phys Anthropol*. 130: 160-166.
[21] Shackelford, L.L. & Trinkaus, E. (2002). Late Pleistocene human femoral diaphyseal curvature. *Am J Phys Anthropol*. 118(4): 359-70.
[22] De Groote, I.E.P.M. (2008). A Comprehensive Analysis of Long Bone Curvature in Neanderthals and Modern Humans Using 3D morphometrics. University College London. PhD Dissertation.
[23] Ruff, C.B. (2000). Body size, body shape, and long bone strength in modern humans. *J Hum Evol*. 38(2): 269-290.
[24] Churchill, S.E. (1994). Human upper body evolution in the Eurasian Later Pleistocene. University of New Mexico. Albuquerque. PhD Dissertation.
[25] Churchill, S.E. (1996). Particulate versus integrated evolution of the upper body in Late Pleistocene humans: A test of two models. *Am J Phys Anthropol*. 100(4): 559-583.
[26] Weaver, T.D. (2003). The shape of the Neandertal femur is primarily the consequence of a hyperpolar body form. *PNAS*. 100(12): 6926-6929.
[27] Stock, J.T. (2006). Hunter-gatherer postcranial robusticity relative to patterns of mobility, climatic adaptation, and selection for tissue economy. *Am J Phys Anthropol*. 131(2): 194-204.
[28] Ruff, C.B. (2000). *Op. cit.*
[29] Stock, J.T. (2006). *Op. cit.*
[30] Rhodes, J.A. & Knusel, C.J. (2005). Activity-related skeletal change in medieval humeri: cross-sectional and architectural alterations. *Am J Phys Anthropol*. 128(3): 536-46.
[31] De Groote, I.E.P.M. (2008). *Op. cit.*
[32] Stock, J.T. & Shaw, C.N. (2007). Which measures of diaphyseal robusticity are robust? A comparison of external methods of quantifying the strength of long bone diaphyses to cross-sectional geometric properties. *Am J Phys Anthropol*. 134(3): 412-23.
[33] Bertram, J.E. & Swartz, S.M. (1991). The "law of bone transformation": a case of crying Wolff? *Biol Rev Camb Philos Soc*. 66: 245-273.
[34] Ruff, C.B. (2008). Biomechanical analyses of archaeological human skeletal samples. In: Katzenburg, M.A. & Saunders, S.R., (editors). *Biological Anthropology of the Human Skeleton*, 2nd Ed. New York: John Wiley and Sons, Inc. p.183-206.
[35] Ruff, C.B., Trinkaus, E., & Holliday, T.W. (1997). Body mass and encephalization in Pleistocene Homo. *Nature* 387(6629): 173-6.
[36] Feldesman, M.R. & Fountain, R.L. (1996). "Race" Specificity and the Femur/Stature Ratio. *Am J Phys Anthropol*. 100: 207-224.
[37] Stock, J.T. (2006). *Op. cit.*
[38] Stock, J.T. (2006). *Ibid.*
[39] Ruff, C.B. (1994). *Op.cit.*
[40] This research was supported by funding from the Natural Environment Research Council, UK. I thank Mercedes Okumura for access to the skeletal material in the Duckworth Collection, University of Cambridge, and my PhD supervisor, Jay Stock, for discussions, ideas, and advice.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

HOW DO WE UNDERSTAND LEONTES? : COGNITION AND CHARACTERISATION IN *THE WINTER'S TALE*

Emma FIRESTONE

PhD candidate, University of Cambridge, Cambridge, United Kingdom
ef262@cam.ac.uk or emma.firestone@gmail.com

Keywords: Shakespeare, Dramatic Romance, Emotion, Memory, Mirror neurons, Empathy, Mimesis

ABSTRACT

William Shakespeare's *The Winter's Tale* is a play that consistently earns its spectators' assent to extremely improbable events. Famously, it defies the laws of physics to represent the (simultaneous) coming-to-life of a marble statue and return of a presumed-dead Queen to the ranks of the living. It also effects a feat of persuasion that defies, just as surely, the laws of ethics: it celebrates the restoration to happiness of Leontes, a man whose acts of cruelty and heartlessness (committed in the first half of the play) are presented as entirely groundless, with no external cause whatsoever. This paper engages with several key concepts from cognitive psychology and social neuroscience, including attentional priming and the mirror neuron system (the neural basis of empathy), in an effort to account for how *The Winter's Tale* secures and maintains this counter-intuitive affective rapport—so difficult to explain, yet so crucial to its aesthetic coherence and success as dramatic Romance.

ON THE PERSISTENCE OF FIRST IMPRESSIONS

A quarter-century ago, the Renaissance literature scholar Stephen Booth described the ability of the mind absorbed in a familiar narrative to retrace its journey from ignorance to surprise—to '*re-capitulate its original responses*' automatically, as these responses are solicited in time—as a crucial mechanism underlying the pleasures of re-reading, repeat theatre-going, and hearing jokes that one already knows. In Booth's intuitive and suggestive account, the successful work of art provokes the same sequence of personalized responses from moment to moment, even when the individual's foreknowledge tells against the validity of these responses with respect to the work as a whole. Thus, in returning to these works, one inevitably, but irrationally,

> '*re-experiences one's ignorance—and re-exercises one's suspicions about Tulkinghorn's murderer,…is delighted that Natasha should once again have had the wisdom to accept Pierre, and is as moved as ever at the eternally unexpected action Othello takes when he says, "I took by the throat the circumcised dog/ And smote him – thus."*'[1]

In setting out these observations Booth was, in his own words, issuing a plea: a plea for critics to reconcile their evaluations of works of art with the '*logically inconvenient common knowledge*' ('derived from logically improbable common experience') that the sort of 'knowledge' one exercises within aesthetic experience can seldom be called logical. For logic dictates that all experiences of narrative that succeed the very first should be radically different from the first: that the spectator who has already seen *King Lear* no longer entertains hopes for Cordelia's triumphant ascent to power and is no longer stunned on beholding her lifeless body; or that having once witnessed Hermione's supremely gratuitous 'return' from (or as) a Statue to the arms of Leontes, we forevermore sit in blunt neutrality as we take in *The Winter's Tale*'s final scene. And of course, this is not the case. Personal experience surely confirms that we consistently experience chronological artwork (lines of verse, whole poems, narratives) chronologically, retaining our capacity for automatic and spontaneous response, thence for renewed joy and reissued pain at each viewing. In ordinary life, the mind, having observed repeatedly that the feelings

ISBN 978-0-9566139-1-2

summoned within a particular context are in fact 'incorrect' or falsely predictive, begins eventually to 'override' those feelings, re-'marking' the occasion with feelings that anticipate the future more accurately.[2] It's through this kind of process, for instance, that most children come to correct the anxiety they experience when standing on top of a playground slide in light of knowledge that they are unlikely to come to harm, eventually replacing their feelings of fright with feelings of glee. Yet when our minds are 'running on' art, our foreknowledge of the character's destinies—or indeed the guesses that our knowledge of genre inclines us to make about these destinies—appears incapable of mitigating what we feel in progress.[3] We never quite start at the beginning; and yet, we never learn.

WHAT THE BODY KNOWS: Emotional memories and narrative progress

Towards the end of his discussion Booth writes, somewhat wistfully: '*What matters much more than why we are capable of such complicated re-experiences is establishing the fact that we* do *regularly undergo such re-experiences. And that fact, unfortunately, is one that cannot be established.*'[4] This is the kind of statement that I seem to have been stumbling across everywhere, in which literary criticism appears practically to cry out for cognitive science to substantiate its bolder intuitions. For research into cognition now *can* establish with reasonable certainty that the phenomenon Booth describes—where the mind 're-experiences' the representational and emotional content associated with sensory or lexical stimuli in familiar contexts, artistic or not—is very real indeed. This mandatory, pre-rational process goes under different names in the cognitive literature addressing the interrelation of narrative, memory, and affective response,[5] but theoretical accounts of the basic steps involved are consistent enough. Essentially, discrete cognitive contents encountered in narratives will 'prime' select personal memories—derived from our own experience, or from other works of fiction, or wherever—that are pertinent to those contents in some way. These representational memories are connected to corollary 'emotional memories' that were 'stored' alongside them and re-activate alongside them too. The process of retrieving the memory, that is, involves the prefrontal cortex in consort with the amygdala engaging the body to re-create the somatic profile that attended the original (remembered) experience; to re-boot the emotion, as it were. (Here I employ Antonio Damasio's influential definition of '*emotion*' as '*a collection of changes in body state that are induced in myriad organs by nerve cell terminals, under the control of a dedicated brain system, which is responding to the contents of thoughts relative to a particular entity or event.*'[6]) The fact that the faces of persons of exceptional value to us (intimates, enemies, offspring) induce the same emotional states each time we see them is a familiar instance of this process. Once the emotion in question has been activated, furthermore, it continues to prime additional salient memories as well as focus the attention on content compatible with the emotion—thereby both sustaining itself and working to direct and delimit cognition in general. Booth's sense that the cognitive contents in works of art issue '*promissory signals*' that create '*expectations*', i.e. that one's present experience (construed in both representational and emotional terms) will mimic the representations proposed by primed memories, quite nicely captures the essence of cognitive science's insights.

It seems to me that insights of this nature propose several axioms of immediate and obvious importance to the study of drama and theatre phenomenology, although none as of yet have found much currency in Shakespeare studies. Foremost among these, perhaps, is that the cognitive processes that regulate our understanding of dramatic fictions are really just the same as those that regulate our understanding of anything in our normal, everyday existences. In everyday life, that is, our understanding proceeds temporally and haphazardly, continually affected by and contingent on our progressively primed memories and the emotions to which these give rise. For each one of us, even those inclined by temperament to 'take the long view' (rather than 'sweat the small stuff'), exists in a condition of continuous pre-conscious attention to the present. Each of us possesses a brain that contains a hypothalamus charged with communicating between the sensory cortex and the body, monitoring the hormone and neurotransmitter levels that generate our somatic condition, controlling the fluctuating internal event that is our emotional state. This is a ceaseless process. Concurrent to this primary monitoring of physical changes is a secondary monitoring process that Damasio calls '*feeling*' and defines as '*the experience of the collection of changes in body state in juxtaposition to the mental images that initiated the [changes].*'[7] Those '*mental images*' are the critical agents here: our feelings are not the 'mental' corollaries to our 'physical' emotions, but represent interpretations of our physical states in light of other information (those 'images') collected in real-time or from primed memories. Yet they are, crucially, neurologically tethered to physical (somatic) events and thus rigorously implicated in the contingencies of emotional/affective condition. A large body of scientific literature further discusses the productive role of feelings in directing our 'attentional focus,' limiting the schemes of assessment

ISBN 978-0-9566139-1-2

and action available to us at that time. All this is to suggest how extensively and undeniably our feelings, and the thoughts they condition, are bound to the present and bound to the body, however unconscious we are of the extent of their influence. Hence, to return to dramatic fictions, our foreknowledge of a sympathetic character's happy fate (say, Helena's in All's Well that Ends Well) will offer little comfort when the feelings provoked by empathy for her plight incline us insistently elsewhere. Just so, in ordinary life, our knowledge that an enticing holiday is coming up next week rarely suffices to mitigate the anxiety of struggling to finish a conference paper.

IMPLICATIONS FOR THE CONCEPT OF DRAMATIC CHARACTER: The case of *The Winter's Tale*

By pursuing these and other insights into the workings of emotional memory, I believe that we can begin to make clear statements about how particular plays in performance make themselves understood; about how, that is, the plays engage and manipulate our emotional responses so as to direct and circumscribe our understanding of what happens in them, imposing conditions on our faculties of response and judgment. I want to propose one condition in particular that might sound tediously self-evident, until one realizes how often and egregiously critics tend to ignore it. This is: that our affective responses to the chronologically earlier content of performed dramatic fictions impose strong affective and interpretive pressures on our reception of chronologically later content—far stronger pressures, in some cases, than critics are inclined to suggest. To some extent, this proposal tests the utility of Booth's assertions (and the cognitive theories they intuit) when applied to the spectator's synchronic experience of dramatic narrative, as opposed to diachronic experience of repeat theatre-going. I am particularly interested in how such an argument might illuminate Shakespeare's strategies of characterization. For just as the spectator engrossed in narrative is impelled through emotion to anticipate outcomes that she knows to be incorrect, so too, I suggest, the spectator's range of feelings towards a particular character over the course of a drama might be (to some degree) enthralled to the chronological sequence of her spontaneous responses to that character. Shakespeare's method of 'constructing' a character might fruitfully be understood as creatively negotiating the dramatic representation of the character *in light of* his spectator's presumed emotional relationship to that character. Taking it as given that our minds *do* recapitulate their original responses to reiterated contents, including characters previously encountered within a play, we might therefore see these original responses as forming an automatic, largely non-conscious foundation for our subsequent assessments of the character—an affective continuity that coexists in productive tension with the temporal dramatic progress of the role.

To recap, what I am attempting to do here is develop a cognition-conscious concept of Shakespearean character that considers the spectator's immediate emotional response to a character as a comparatively uniform, and highly influential, component of the spectator's understanding. This component can also be understood as the character's tendency to engage particular, innate emotional triggers that humans hold in common.[8] The work of 'characterization' is therefore to be understood as a dynamic process of negotiating this primary, emotionally engendered component with subsequent representations of the character—representations that often run contrary to the 'expectations' raised by the original responses. My chief interest is in illuminating the murky concept of theatrical understanding, potentially clarifying what is meant when we say that we understand a character or alternately, that a character eludes our understanding. One working hypothesis is that our perception of 'failing to understand' a character may point to primary emotional responses that run contrary to sequent representations of that character; in Booth's phrase, that '*beckon elsewhere*'.

To begin to substantiate this concept, however, I want now to consider an instance in the Shakespeare canon in which emotional response works 'invisibly' to *promote* rather than frustrate a sense of understanding: that is, where we have the sensation of empathizing with a character whose behavior demonstrably makes no sense. I believe that this is what occurs in the case of Leontes, a character who enjoys a curiously untroubled and sympathetic affective rapport with audiences, given that his 'affection' or emotional condition is so insistently 'diseased', his actions transparently deplorable. That this should be so owes, I submit, to the affective force of our first impressions of a single character, the character of Hermione. I will develop this claim in the space I have left.

With the exception of prince Hamlet, Leontes is distinguished among Shakespeare's characters for spending the majority of his scenes behaving in a manner that every onstage commentator, and there are many of them, deems

ISBN 978-0-9566139-1-2

a wild aberration from his usual self. He is a character who appears before us, as it were, out of character. For spectators this is a strikingly curious state of affairs: just off the street and barely settled in our seats, on what grounds can we be made to accept that a character whom we have only just met is not acting himself? Significantly contributing to the challenge to spectatorial empathy is the fact that Leontes' evident emotional state has (notoriously) no demonstrable source. It lacks the basis in ocular proof, in the rational if falsely motivated conclusions behind the suspicions of Posthumus or Othello. Despite all this, there are unusually high narrative stakes associated with securing empathy for this character: for in due course, after observing three acts full of egregious and unfounded wrongdoing, we will be solicited in Act Five to assent to his restoration to happiness. And our assent, thus granted, will precipitate a concluding scene of which A.D. Nuttall has written: '*Most of us have never experienced anything so wonderful and never will, but Shakespeare shows us what it might be like if we did.*'[9]

The production history of *The Winter's Tale* is long and successful, so it should not in itself surprise that spectators consistently do rise to the empathetic challenge posed by the out-of-character character of the king. Consistently, they recognize Leontes' behavior in Scene 1.2 as aberrant, as portending something that is not quite moral guilt, and therefore merits forgiveness. But *how* spectators reach this understanding is a complex matter indeed. What I believe happens is that spectators implicitly apprehend a latent positive wonder within Leontes' jealousy—that they assume the presence of love beneath his ardent negation of love. More germane to my present argument, this feat of empathetic intelligence, which permits spectators to comprehend Leontes' experience of suspicion without condoning his error in judgment, is facilitated by the spectators' own local, concurrent feelings of love. And the source of this love is the character of Hermione.[10] I am proposing that the intelligibility of *The Winter's Tale* in performance frankly depends on the character of Hermione's ability to evoke feelings of love in spectators. Love, inspired by the figure of Hermione and concentrated around her onstage person, functions in *The Winter's Tale* as an assumed affective response (assumed, that is, by Shakespeare), with a structural and interpretive function as integral to the play's overall coherence as the ordering of its scenes.

A SOCIAL NEUROSCIENCE INTERVENTION: Mirror neurons and theatrical empathy

Turning briefly to what transpires at the neural level to effect our understanding of Leontes' groundless transformation, I will briefly evoke the science of 'mirror' neurons and their well-known role in facilitating interpersonal understanding at the level of embodied action. Mirror neurons encode templates for specific actions and store the sensory-motor information associated with them, enabling us not only to recognize actions intuitively but internally to simulate the actions ourselves.[11] Research on the mirror neuron system (MNS) thus illuminates the cognitive processes behind comprehending and responding to others' actions, emphasizing the role of 'embodied simulation': a pre-rational mechanism that simulates the action or emotion observed, generates an image of its goal or intention based on inference and environmental cues, and imaginatively attributes this goal to the action.[12] Hence when we observe behaviors in others, we also internally simulate the states associated with the actions and emotions we observe, 'as if' we were experiencing the same thing.[13] Against conventional theories that posit interpersonal understanding as based on rational efforts to 'step into somebody else's shoes,' then, embodied simulation proposes understanding others as an automatic, unconscious process. The MNS therefore ensures that our perception of Leontes watching Hermione in *The Winter's Tale* Scene 1.2 inclines us automatically to imagine the quality of attention i.e. the emotions governing that watching. (The fact that a dramatic character's emotional state is always imaginary is thus of little consequence; the neural processes engaged in understanding are the same, whether the object of that understanding is real or fictional.) I suggest that we *do* apprehend something of this character's mental/intentional state—his fascination with Hermione, his recognition of her charms and how she enacts them before himself and Polixenes—both because these qualities accord with our personal experience of watching in general, and because they are legitimated instantly by our experience of admiring the queen.

This last point is crucial. In light of the process discussed above wherein particular emotional responses and the expectations thus engendered influence our understanding, the chief 'work' of Scene 1.2 is to render the spectator's affection for Hermione, thence his conviction of her virtue and chastity, absolutely unshakeable. I am less concerned to articulate precisely how this happens than to suggest it is essential for the play's coherence that it *does* happen. Though as to the former, Anthony Dawson's idea that Hermione's attractiveness depends on Shakespeare's yoking her character to the intensely, affectively charged notion of 'grace' is highly suggestive, in large part for its clear relevance to the workings of emotional memory as discussed above.[14] And certainly the

ISBN 978-0-9566139-1-2

notion of 'grace' appears an indispensable feature of the character as performed. Hermione's grace is acknowledged by practically everyone in Sicilia and resounds significantly through representations of her imprisonment (of which she declares, 'this action I now go on / Is for my better grace' [2.1.112]) and trial.[15] To call this character 'seductive' cannot be scandalous, for spectators, in their desire to see her honour recognized and reputation restored, can fairly be said to have fallen in love with her; or perhaps more accurately, to harbor feelings of love that are triggered by the character and concentrated around her. That Shakespeare has *assumed* this response is evidenced by his inclusion of multiple scenes that presume, and represent, unconditional support for Hermione against her husband. Among these, none is more touching than the stage-picture of her retreat to prison, where her natural authority contrasts with Leontes' undignified ranting. Attended by devoted ladies, the queen's grace and honor are so evident as to summon boldness from the choric Lord ('*Beseech your Highness call the queen again*' [2.1.126]), who cannot bear her to be slandered or even to leave the stage.

Importantly, too, the spectator's susceptibility to Hermione's gracious presence in 1.2 implicitly reenacts feelings attributed in the scene to the younger Leontes, won by Hermione's charms despite their inevitable connections to the sphere of adult sexuality. Here, Damasio's distinction between emotions and feelings provides an important interpretive tool. The inferences that the MNS prepares us to make about other's feelings (the content of their emotions) are inferences only, guesses made by combining experiential knowledge with such immediate biasing factors as our local physical conditions. They can be wrong; they often are. I wrote above that the onset of Leontes' *tremor cordis* has no demonstrable source in the world; that as his address to this 'affection' puts it, '*With what's unreal thou co-active art / And fellow'st nothing*' [1.2.141-2]. But *crediting* Leontes' dancing heart is not only possible, but easy. Playgoers need do nothing more than mark their own emotional response to the character of his queen: their feelings of love, source of their convictions of her virtue. This experience of conviction is then enlisted to meet the empathetic challenge detailed earlier, enabling us to comprehend Leontes' sudden transformation by illogically equating our experience of love with his experience of suspicion. That is, the spontaneous manner in which the rapt audience apprehends Hermione's grace is analogous to the spontaneous manner in which the rapt Leontes develops his jealousy. And our experience of love helps us to credit his jealousy and ultimately, to correct it.

To summarize, the MNS enables us to simulate and comprehend Leontes' affection but not his peculiar interpretation of that affection. Leontes determines his affection to portend jealousy; but as a physical reaction, it has a near-identical structure to the positive, creative feelings of wonder and love. The structural congruity of the apparently incongruous feelings of jealousy and love indeed provides *The Winter's Tale* with one of its governing sources of structural and thematic coherence: for the opposition between jealousy and love is one of many—including guilt and innocence and even, via the uncanny and spellbinding Statue of 5.3, life and death—that play effects to expose as variant interpretations of a single condition. Put another way, the playgoer's experience of watching Leontes' performance of jealousy while strongly influenced by its own feelings of love emphasizes the paradox that inheres in Hermione's grace. To watch 1.2 is to see admiration cast as suspicion, innocence cast as guilt, but although only one interpretation is consistent with the audience's affective experience, both are valid within the scene's own discursive terms. Many critics note that Hermione's and Polixenes' conversation is laced with language suggesting Fallenness and post-lapsarian regret. However deftly the actors handle this language, however transparently inappropriate a 'sinful' interpretation of the scene, there is inherent in their speech enough allusive force and imagistic potential to incline Leontes toward a negative interpretation of his emotional response to the action—to support a 'reading' of his feeling that gives suspicion priority over admiration.

CONCLUSION: Why good actors matter

Thus spectators come to 'understand' Leontes' dancing heart as an emotional symptom prompted by the very same stimulus that grounds their own emotional relationship to the scene: wonder at Hermione's manifest grace. This analysis has the virtue of conferring affective coherence on the character of the king, establishing precedent for his wondering and deferential persona in the Statue scene and indeed throughout Act Five. More importantly, perhaps, it provides support for the idea that Leontes is moved to interpret his emotion erroneously under the cumulative pressure of misleading environmental and emotional cues. The spectator, charmed by Hermione's grace and therefore convinced of her virtue, has the upper hand on Leontes' interpretation and rightfully so: the play's generic and affective terms do not license another interpretation. But as I have variously insinuated, any performer of Hermione who is too suggestive, too demonstrative about the queen's carnal

ISBN 978-0-9566139-1-2

knowledge, or simply not charming enough can make *The Winter's Tale* considerably less intelligible. Cognition-oriented performance criticism, in taking emotional response seriously, must be prepared to make qualitative judgments as to which kinds of actors are best equipped to do their characters justice.

REFERENCES

[1] Stephen Booth, 'Appendix 1: On the Persistence of First Impressions', in King Lear*,* Macbeth*, Indefinition, and Tragedy* (New Haven: Yale University Press, 1983), 121-4.
[2] For the role of 'somatic markers' in facilitating response and decision-making, see Antonio Damasio, *Descartes' Error* (New York: Harper Perennial, 1995), especially 173-9.
[3] Phrase borrowed from Keith Oatley's concept of fictional stories and characters as 'simulations of the social world that run on minds, [much] as simulations of weather systems run on computers.' Oatley's discussion of character in relation to emotions chimes with Booth's sense that artistic experience involves exercising a distinctive kind of knowledge: 'As we run literary simulations in our minds, we not only *imagine the characters*, we *experience the emotions and hence the urgency of their human vicissitudes and dilemmas*, and we are enabled…to create mental models of ourselves…'. See Oatley, 'Simulation of Substance and Shadow', *College Literature*, 33.1 (Winter 2006), 17-18.
[4] Booth, 124.
[5] Within cognitive theory these terms include 'priming' and 'activation of emotional nodes'; a lucid survey of different, but compatible accounts of these processes is in Patrick Colm Hogan, *The Mind and its Stories: Narrative Universals and Human Emotion* (Cambridge, CUP, 2003), especially 54-71; also Hogan, *Cognitive Science, Literature, and the Arts* (New York: Routledge, 2003), 155-65.
[6] See Hogan, *Cognitive Science*, 2003, 182-84; Antonio Damasio, *Descartes' Error* (New York: Harper Perennial, 1995), 139.
[7] Damasio, 145.
[8] Hogan, *Cognitive Science*, 185.
[9] A.D. Nuttall, *The Winter's Tale* (London: Edward Arnold, 1966), 58.
[10] My subject here is the play in performance, hence my remarks pertain to Hermione as 'incarnated' in the body of an actor. I assume that Shakespeare's plays pursue the traditional affective goal of theatre, that the actor's performance mediates the spectator's understanding of a character's feelings; at this point, distinctions between Hermione and her performer obviously begin to collapse. It strikes me that the potential of applying affective cognition to Shakespeare studies includes the ability of such research to articulate the cognitive advantages of performance—to make strong claims as to what a spectator 'gains' from experiencing the play performed live.
[11] Vittorio Gallese et al., 'Intentional Attunement', *Journal of the American Psychoanalytic Association*, 55 (2007), 131-176; M. Iacoboni et al., 'Grasping the Intentions of Others with One's Own Mirror Neuron System', *PLOS Biology*, 3 (2005), 1-7.
[12] Gallese, 'Embodied Simulation', *Phenomenology and the Cognitive Sciences*, 4 (2005), 23–48.
[13] Gallese, 2005, 27.
[14] 'Shakespeare and Secular Performance', in *Shakespeare and the Cultures of Performance* (Ashgate: Burlington, VT, 2008), 92-95.
[15] e.g. Camillo, "To satisfy your Highness and the entreaties / Of our most gracious mistress" (1.2.232-3); Polixenes, "Good expedition be my friend, and comfort / The gracious queen" (1.2.458-9); Lord, 2.1.129-33; Paulina, "How fares our gracious lady?", "a gracious innocent soul, more free than he is jealous" (2.2.19, 29-30); Dion, anticipating the oracle's verdict, "gracious be the issue!" (3.1.22).

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

MENTAL HEALTH COURTS

Kelly FRAILING

PhD Student, University of Cambridge Institute of Criminology, Cambridge, UK
kf277@cam.ac.uk

Keywords: Mental health courts, Referrals, Outcomes, Sanctions, Perceptions, Controversies

ABSTRACT

Mental health courts (MHCs) are a relatively new phenomenon in the United States. One of a variety of specialty courts that aim to ameliorate the problems which can bring offenders into initial and repeated contact with the criminal justice system, mental health courts divert willing and eligible mentally ill offenders from incarceration toward mental health and related treatment, which can include psychiatric medication, counselling, substance abuse treatment, job training and housing acquisition. Following a description of the court and its procedures, this paper examines various aspects of the Washoe County, Nevada mental health court, including referrals, outcomes, the use of sanctions for noncompliance and participant and staff perceptions. Evidence gathered thus far suggests that the Washoe County mental health court is accepting some of those in need of treatment, including those with criminal histories, is successful at reducing criminal justice and substance abuse involvement for its mentally ill participants, that sanctions, including jail, are ordered for noncompliance with treatment and that it meets with the approval of its staff and participants. However, these courts are not without their controversies, including the use of certain sanctions, appropriateness of rejecting eligible referees and the expansion of the judge's role to include that of mental health court team member and these issues are examined as well.

INTRODUCTION

Thanks in no small part to the conservatively oriented criminal justice system that has characterized the United States in recent decades, the criminal court system has long been overburdened with cases and consistently ill equipped to handle the complex issues offenders can bring to bear. These issues, which can result in initial and repeated contact with the criminal justice system, include but are not limited to drug addition, mental illness, homelessness and domestic violence involvement. In response, a number of jurisdictions have developed problem solving courts that aim to more effectively address these chronic problems by diverting willing and eligible offenders toward treatment instead of incarceration, with the goals of reducing jail overcrowding and recidivism and improving offenders' quality of life. One such problem solving court is the mental health court (MHC). There are over 250 mental health courts in the United States today.[1]

Following a review of the relevant mental health court literature and a description of the Washoe County mental health court, this paper examines referrals, outcomes, sanctions ordered and perceptions of the court and discusses some attendant controversies.

Literature review

Research on mental health courts is still in its early years. Only two studies to date have focused on referrals to the court. The first of these found that referees' characteristics significantly related to rejection from mental health diversion included a history of felony convictions, a current charge of a crime against a person and being male. Age was affected by gender so that overall, younger women's and older men's referrals were more likely to be accepted.[2] The second referrals study found that 30 percent of all rejections from seven mental health courts were due to mental health status and about 20 percent were due to past or current criminal charges. Those

ISBN 978-0-9566139-1-2

diagnosed with schizophrenia or bipolar disorder were more likely to be accepted into the courts. There were no significant findings for age, race or gender in this study.[3]

More research has been done on criminal justice outcomes for mental health court participants. A number of studies have found that participation in the mental health court is associated with reduced recidivism. Mental health court participants had fewer arrests than before their enrolment in the court,[4, 5, 6, 7] they had fewer criminal charges during mental health court than before their enrolment,[8, 9] they had a decrease in jail days,[10, 11, 12, 13] they were less likely to commit crimes and violent crimes after participation as compared to a control group[14] and the positive effects of mental health court participation on recidivism were retained for years after graduation.[15] Some studies have also examined substance abuse outcomes for mental health court participants and have found that mental health court participation is associated with reduced substance abuse.[16, 17, 18, 19]

Mental health courts use sanctions when participants are noncompliant with court mandated treatment and a few studies have examined the use of sanctions at different courts. Early research on eight mental health courts found that jail was rarely if ever used as a sanction for noncompliance in seven of the courts surveyed. The one court at which it was more regularly used accepted referrals with a felony index charge.[20] More recent research found that the newer generation courts used jail as a sanction for noncompliance more often than the older ones, likely because the newer ones accepted those referees with a current felony charge.[21] A survey of 90 mental health courts revealed that the use of jail as a sanction for noncompliance varied along a rough bell curve, with a few courts never using jail, many courts using jail rarely, occasionally or sometimes and a few courts using jail in over half of cases.[22]

Participant and staff perceptions of the court have also garnered some research interest. One study found that mental health court participants perceive little coercion and greater procedural justice than do traditional court participants.[23] Another study found that both participants and staff support the mental health court and consider it effective.[24]

The Washoe County mental health court

The present study was conducted at the Washoe County, Nevada, United States mental health court, which is located in Reno, the county's largest city. The court was established in 2001 and targets those with diagnoses of schizophrenia, bipolar disorder and major depression with a variety of criminal charges (misdemeanours and felonies) in an effort to reduce jail overcrowding and recidivism. About 200 people are participating in the Washoe County mental health court on any given day.[25]

The Washoe County mental health court requires a plea of guilty or *nolo contendere* or a finding of guilty for referees to be accepted. There is a minimum of one year of participation in the mental health court and a more serious charge may warrant a longer enrolment. There are two mental health court judges who rotate with one another. Both have been appointed to the mental health court by the Nevada State Supreme Court. The mental health court team is comprised of the current judge, the defence attorney, court staff and mental health clinicians and service providers. The mental health court team works with each participant to develop an individualized plan of care to help them achieve independence and an improved quality of life. The plan of care can include services provided by court personnel, such as drug testing and regular court dates from every one to four weeks, as well as those provided by mental health clinicians and service coordinators, such as outpatient psychiatric and psychological treatment, substance abuse treatment, job acquisition and supported living.[26]

METHODOLOGY

To examine referrals, outcomes, sanctions and perceptions at the Washoe County mental health court, a record review, surveys and observations were employed. Both examining a number of aspects of the court in a single study and using mixed methods to do so are rarities in the mental health court literature. Demographic, mental health and criminal characteristics of referrals to the court from 2006 through 2009 were examined with logistic regression to determine which were associated with acceptance and rejection. Not included in the regression analysis was referees' criminal history information, so these data were examined separately.

ISBN 978-0-9566139-1-2

Several outcomes were examined in this study. The number of jail days served by 2007, 2008 and 2009 participants and graduates (those who successfully complete mental health court) were examined for the year before enrolment, the year of enrolment in the mental health court and the year after graduation. These jail days were compared to those served by a comparison group, comprised of those referees who were accepted into mental health court but who had another disposition of their cases that prevented them from participating and those who were accepted but chose not to participate. Psychiatric hospitalization days were also examined for 2007, 2008 and 2009 participants and graduates in the year before enrolment, the year of enrolment and the year after graduation. Drug and alcohol test results were measured by counting positive tests for each current participant over the course of a year and noting when during the enrolment they occurred. Each instance of the use of community service and of jail as a sanction was counted over the course of a year and the ordering of sanctions by the judge was observed in the mental health court sessions.

Participants' and staff's perceptions of the court were measured using original surveys administered within two weeks to newly enrolled and graduating participants and to staff at two points one year apart. The participant surveys were written to ascertain perceptions of fairness, voice, coercion and satisfaction at enrolment and graduation. The staff surveys were written to ascertain staff's perceptions of the court, as well as their perceptions of participants' perceptions of the court. The surveys were designed so that participants' perceptions could be compared at enrolment and at graduation and so that participants' and staff's perceptions could be compared to one another.

RESULTS

Referrals

Demographic, mental health and criminal data were gathered on 1,220 referrals to the mental health court from 2006 through 2009. Each level of the seven categorical variables (age, race, gender, referral source, type of index charge, seriousness of index charge and mental health problem) was subjected to logistic regression to determine which significantly predicted acceptance into and rejected from the mental health court. A crime against a person, a crime against the community and being male were all significantly predictive of rejection from the mental health court at the .05 level. Misdemeanour, gross misdemeanour and felony as the seriousness of the index crime, as well as the mental health problems of thought disorder, mood disorder, anxiety disorder and mental retardation were all significantly predictive of acceptance into the mental health court. Age, race and referral source were not predictive of acceptance into or rejection from the mental health court. The odds ratios indicated that some levels were stronger predictors than others. Those with retardation, with a thought disorder, with a mood disorder, charged with a felony, with an anxiety disorder, charged with a misdemeanour and charged with a gross misdemeanour were respectively 72, 12, 12, eight, seven, seven and six times more likely to be accepted into the mental health court that those without these characteristics. The odds ratios for significant predictors of rejection were not nearly as high. Those with a crime against a person were just .23 times more likely to be rejected, those with a crime against the community .29 times more likely to be rejected and males .63 times more likely to be rejected than those without this characteristic.

Over half of the referrals to the mental health court during this time period (N=678) were rejected. The most common reason for rejection was other disposition of the case (N=287), meaning the referee was sentenced to time served, to prison, etc and ineligible to participate in the mental health court. The second most common reason for rejection was lack of severe mental illness or treatment history (N=201) and the third most common reason for rejection was criminality, either nature of current crime or criminal history (N=117). Because criminal history information was not included in the logistic regression predicting acceptance or rejection into the court, these data were analyzed separately. Data were available for 58 of the 84 referees rejected for the court-stated reason of seriousness of criminal history. These 58 referees were charged with a total of 518 crimes in the adult lifetime, with an average of 8.93 per person. With some exceptions, there seemed to be a trend evident in which the referees rejected for criminal history reasons either had a long criminal history with relatively minor charges or a short criminal history with more serious charges. The most serious crimes committed by those rejected for criminal history reasons were violent crimes against people, i.e. manslaughter, assault with a deadly weapon, rape, robbery, etc.

ISBN 978-0-9566139-1-2

Outcomes

Chi square test were performed on the mental health court participants and graduates (N=146) and on the comparison group (N=238) to determine their suitability for comparison. The tests revealed that the comparison group had a higher proportion of males than the mental health court participants and graduates group (χ^2 (df=1, n=380)=5.5, p=.025, phi=-.120) and the comparison group had a lower proportion of drug crimes and a higher proportion of crimes against the community than did the mental health court participants and graduates (χ^2 (df=5, n=380)=25.6, p=.000, phi=-.260). None of the other comparisons, between age, race, referral source and mental health problem, revealed significant differences.

Though the groups were not identical and though the comparison group was not randomly selected from the population at large (and hence is not referred to as a control group), a comparison between the two was still valid, as the comparison group was comprised of those people accepted into but not enrolled in the mental health court and had they not had another disposition of their cases or elected not to participate, they would have been mental health court participants. Tests of significant difference were conducted on the mental health court participants' and graduates' and comparison group's jail days and psychiatric hospitalization days for the year before referral, the year after referral (the year of enrolment for participants and graduates) and the year after graduation. Table 1 shows the results of these tests.

Table 1. Comparison of comparison group (N=238) and participant and graduate group (N=146) on jail days and psychiatric hospitalization days

JAIL DAYS	Mean	N	St Dev	t	df	p< (2-tailed)
CG-before	56.54	238	73.67	-.01	312	.991
MHCPG-before*	56.62	145	71.29			
CG-after referral	134.61	238	150.23	12.34	256.61	.000
MHCPG-enrol*	11.91	146	24.13			
CG-before	56.54	238	73.67	-8.32	237	.000
CG-after	134.61	238	150.23			
MCHPG-before	56.62	145	71.29	7.57	144	.000
MCHPG-enrol	11.99	145	24.19			
MHCPG-before	60.09	86	72.85	7.33	85	.000
MHCPG-after	5.79	86	25.84			
MHCPG-enrol	15.73	86	29.46	2.50	85	.014
MHCPG-after	5.79	86	25.84			
CG-after referral*	134.61	238	150.23	12.72	272.3	.000
MHCPG-after*	5.79	86	25.84			
PSYCH HOSP DAYS	**Mean**	**N**	**St Dev**	**t**	**df**	**p< (2-tailed)**
MHCPG-before	2.42	107	8.13	2.18	106	.031
MHCPG-after	.66	107	3.54			

Where CG=comparison group, MHCPG=mental health court participants and graduates, before=the year before referral, after referral=the year after referral for the comparison group, enrol=the year of enrolment for mental health court participants and graduates and after=the year after graduation and where * indicates an independent samples t-test was used and equal variances not assumed (all others were paired samples t-tests).

While the comparison group and mental health court participants and graduates served a statistically similar number of jail days in Washoe County in the year before referral, the mental health court participants and graduates served significantly fewer jail days than did the comparison group during the year after referral (the

ISBN 978-0-9566139-1-2

year of enrolment) and during the year after graduation. Moreover, the mental health court participants and graduates served significantly fewer jail days in Washoe County during enrolment and after graduation as compared to the year before their referrals. It is notable that a majority of Washoe County mental health court participants were chronic recidivists before their participation, with 59 percent having two or more charges in the adult lifetime. The comparison group actually served significantly more jail days in the year after referral than in the year before. The mental health court participants and graduates also spent significantly fewer days in the psychiatric hospital in the year after their graduation as compared to the year before their referral.

The frequency and timing of drug and alcohol tests were counted for current mental health court were counted for one year. Current participants received 1,170 drug tests, 203 (17.4 percent) of which were positive. More than half of the 203 positive drug tests (112) occurred within the first four months of enrolment in the mental health court. Almost 70 percent (142) occurred in the first six months of enrolment. Current participants also received 3,040 alcohol tests, 64 (2.1 percent) of which were positive. More than half of the 64 positive drug tests (35) occurred within the first four months of enrolment in the mental health court. Over 65 percent (42) occurred in the first six months of enrolment.

Sanctions

The type of sanction ordered for current participants as recorded in the database maintained by mental health court staff were counted for one year. There were 401 sanctions given to participants for noncompliance; 200 (49.7 percent) were community service and 201 (50.3 percent) were jail days. These are the only types of sanctions explicitly recorded in the database. As seen below, the court employs other types of sanctions where warranted.

Observations of the mental health court sessions also revealed trends in the ordering of sanctions. Both mental health court judges as well as the one substitute judge observed did not order sanctions very often, issuing praise and encouragement to participants as they came before him far more often. Table 2 illustrates this trend.

Table 2. Judges' remarks during Washoe County MHC sessions

Date	Praise	Encouragement	Sanction
December, 2008 (Judge 3)	21	24	9
April, 2009 (Judge 2)	35	13	4
July, 2009 (Judge 1)	31	25	6
TOTALS	**87**	**62**	**19**

Praise included remarks such as: '*great example to the group*', '*congratulations on doing great*', '*good for you*', '*you're doing great*', '*off to a good start*', '*best week in a while*', '*looking really good*', '*good job*', '*wonderful*', '*another perfect week*', '*you're making us proud*', '*way to go*' and '*you're perfect*'. Encouragement included remarks such as: '*keep up the good work*', '*stay on track*', '*work through it*', '*get serious with the program*', '*you can do it*', '*keep it up*', '*you stick with us and we'll help you*', '*cowboy up*', '*better results from now on*', '*don't use [drugs] this week*', '*keep up the routine*' and '*keep doing what you're doing*'.

Sanctions included orders such as: perfect or else (meaning that the participant had to have a perfect week in terms of court and mental health obligations or he or she would face additional, harsher sanctions), community service for drinking, which is prohibited during mental health court, jail for missing court and mental health service obligations all week without a legitimate excuse and observing all specialty court sessions the next week for not keeping appointments with mental health professionals. When delivering a sanction, the judge spoke more slowly, loudly and in a deeper tenor, as though to convey the gravity of his words, sometimes reviewing the specific reasons for which the participant was receiving the sanction. This was also the case when he threatened a sanction but did not order one.

Perceptions

Surveys were administered by court staff to participants upon enrolment in (N=55, 63 percent response rate reported) and upon graduation from (N=44, 68 percent response rate reported) the mental health court and the

ISBN 978-0-9566139-1-2

responses on these surveys were compared using tests of significant difference. The staff responses from the baseline and one year administration of their surveys (N=14 each, 100 percent response rate) were also compared. Staff responses to questions on participants' perceptions were also compared to participants' responses. Table 3 displays the significant results of these comparisons.

Table 3. Significant differences in Washoe County MHC enrolees (N=55), graduates (N=44) and staff (N=14) survey responses

	Mean	N	St Dev	t	df	p<
Graduation-fairness	6.16	44	1.19	2.28	56	.027
Staff (2nd administration)-participants fairness*	5.25	14	1.63			
Enrolment-fairness	5.45	42	1.60	-2.53	41	.015
Graduation-fairness	6.17	42	1.21			
Enrolment-satisfaction	5.45	42	1.45	-3.80	41	.000
Graduation-satisfaction	6.38	42	.96			

*Independent samples t test, equal variances assumed. The remainder of tests performed were paired t tests.

The graduating participants thought the court was significantly more fair than the staff thought they did. The graduating participants also thought the mental health court was significantly more fair and were significantly more satisfied with it than were newly enrolled participants.

As seen in Table 3, the mean responses on the survey items were generally high on the scale of one to seven where one indicated strong disagreement and seven indicated strong agreement. An examination of responses on all survey items by all participants and staff revealed that none of the mean responses was below five, indicating that participants and staff had generally positive perceptions of the mental health court.

DISCUSSION

Taken together, these results indicate offenders with mental illness who are charged with crimes against people or the community or those with violent criminal histories are unlikely to be permitted to participate in the mental health court. For those who are permitted to participate, namely those with thought or mood disorders and those charged with misdemeanours or nonviolent felonies, the mental health court is an effective intervention associated with reduced recidivism, even for chronic recidivists, as well as reduced intensive psychiatric service use and drug and alcohol use. Mental health courts issue a variety of sanctions, including jail, for noncompliance and the court sessions are non-adversarial in nature. Participants and the judge interact directly and praise and encouragement are issued far more often than sanctions. Both participants and staff have positive perceptions of the mental health court. Other studies have found similar results, but this is the first simultaneous examination of several aspects of the mental health court and it permits elucidation of the court's key features, namely who it serves, how and to what ends.

A broader question is what the mental health court's place is in the local criminal justice system. While the name mental health courts may imply that these courts are available and appropriate for all offenders with mental illness, this is clearly not the case. The court in this study requires a guilty plea to enter, meaning this court is a post-adjudication intervention and participants must accept a new criminal charge in order to enrol. The court is reticent to admit those with violent crimes or criminal histories, meaning the mentally ill with these characteristics are often subject to traditional criminal justice processing. Moreover, otherwise eligible referees who had another disposition of their case were also subject to traditional criminal justice processing. Mental health courts are not a panacea, but they are cognizant of the reality of the criminally involved mentally ill population and they provide mental health and related treatment in the community to this population while using criminal justice means to preserve public safety. Mental health courts are an increasingly popular addition to local

ISBN 978-0-9566139-1-2

criminal justice systems in the United States and, as the current research has shown, are effective interventions for some of those offenders with mental illness without a history of violence.

CONCLUSION

Though the mental health court appears effective and satisfactory, it is not without its controversies, among them the use of certain sanctions, appropriateness of rejecting eligible referees and expansion of the judge's role to include that of mental health court team member.

As seen above, the Washoe County mental health court utilizes sanctions up to and including jail for noncompliance with treatment. Some critics believe the use of jail in the mental health court is inappropriate because the court is supposed to be therapeutic for participants. Jail is not therapeutic and therefore has no place in the mental health court.[27] Other critics believe jail deprives participants of the opportunity to engage in treatment.[28] Observations of the mental health court revealed that jail was not used lightly or capriciously as a sanction, but only in the cases of serious or repeated noncompliance. It was also used proportionally, with a few days of jail time the maximum for the most serious instances of noncompliance. The use of jail is not statutorily limited and its even-handed application in Washoe County is indicative of appropriate discretion on the part of the mental health court judges there. Disallowing jail as a sanction within the mental health court would mean fewer options for the judge to demonstrate to participants that they are indeed responsible for their own choices.[29] The remaining options, absent jail, might not be able to press this point upon some participants, though the effectiveness of jail as a sanction for noncompliance within the mental health court remains an uninvestigated area. Repeated and serious noncompliance can result in removal from the mental health court, but the use of increasingly harsh sanctions, up to and including jail before removal for those struggling to adhere to the court's requirements permits both accountability for noncompliant actions and the opportunity for participants to continue to try to engage with available services.

The above discussion of the rejection of referees for reasons of nature of current crime and criminal history leads inevitably to the question of whether it is appropriate for the mental health court staff to use its discretion to reject those referees with a serious current crime or criminal history who otherwise meet the requirements of court eligibility or whether the court should accept all those who technically qualify and want to enrol. The statutes that determine who technically qualifies for the mental health court do not eliminate anyone on the basis of criminal history, noting only that the district attorney must approve the enrolment of violent offenders.[30] The court's rationale for rejecting referees based on current crime and criminal history is that doing so preserves public safety. It would be highly problematic for the court, not to mention the victim of the crime, if any participant were to commit a violent crime while enrolled so court staff hedges their bets, sometimes exercising their discretion when a referee has a violent current charge or criminal history. Research has shown that the best predictor of future criminal behaviour is past criminal behaviour[31, 32, 33] and that the same is true for the mentally ill.[34, 35] In rejecting current or formerly violent referees, the court staff places the responsibility for their incarceration or other management in the hands of another entity, i.e. the traditional court. As for the ethical dilemma of where or how these referees are going to receive the mental health treatment they need, the jail in Washoe County is able to offer some mental health services, so a rejection from the mental health court based on current crime or criminal history does not necessarily mean the referee will go without treatment, though treatment offered by the jail is less comprehensive than that offered through mental health court. When the court in Washoe County was first established, it, like many mental health courts at the time, accepted a majority of referees with misdemeanour charges. This gradually changed and now there are more enrolees with felony index charges than with misdemeanour index charges. This willingness to, over time, accept those with more serious index charges reflects a confidence in programming on the part of mental health court staff and in the future, the court may become more comfortable with accepting more of those referees with violent current crimes and/or criminal histories. The mentally ill with violent crimes and/or criminal histories who have been treatment resistant in the past may benefit from not only the treatment of mental illness and substance abuse and the supports the mental health court has to offer, but also its structure and accountability. Involvement with treatment is associated with reduced risk of recidivism and violence and insofar as the mental health courts can sustain this involvement of participants, it has potential to be the appropriate setting in which to serve at least some of the mentally ill with violent current crimes or criminal histories.[36]

ISBN 978-0-9566139-1-2

Lastly, the Washoe County judge plays a variety of roles in the mental health court that go beyond the traditional judicial role of impartial arbiter. In his role as mental health court staff group member, he participates in staff meetings, which means he discusses participants with other staff and provides as well as receives input on whether to issue praise, encouragement or sanctions to current participants. In his role as mental health court judge, he issues praise, encouragement and sanctions, up to and including removal for noncompliance. It is important that the judge balance his roles of staff group member and judge and not allow the latter (and the power and finality it bespeaks) to overrun the former. That is, when the judge is acting as a member of the staff, he should act as one, with his input equal to that of the other members. When he is acting as the judge, he must maintain impartiality, especially when issuing sanctions and deciding whether to remove noncompliant participants and his decisions must be based on the factual record and not on bias. He must also be able to interact with participants in a meaningful way, especially when engaging in the non-traditional activities of issuing praise and encouragement, as the positive relationship between the judge and the participant is associated with the positive outcomes noted above.[37] Observations revealed the judges in the Washoe County mental health court seemed comfortable with and competent at striking a balance between the roles of group member and judge, so it is possible for a mental health court judge to play traditional and non-traditional roles simultaneously and effectively.

Using mixed methods to examine a number of different aspects of the mental health court leads to informative results that increase these understanding of who these courts serve, how and to what ends. However, it is imperative that controversial issues inherent in the courts' operation be examined in tandem with results so that potentially troubling issues are equally well understood.

REFERENCES

[1] Almquist, Lauren and Dodd, Elizabeth. (2009). *Mental Health Courts: A Guide to Research-Informed Policy and Practice*. New York: Council of State Governments.
[2] Luskin, Mary Lee. (2001). Who is diverted? Case selection for court-monitored mental health treatment. *Law and Policy, 23*(2): 217-236.
[3] Steadman, Henry J., Redlich, Allison D., Griffin, Patricia, Petrila, John and Monahan, John. (2005). From referral to disposition: Case processing in seven mental health courts. *Behavioral Sciences and the Law, 23*: 215-226.
[4] Christy, Annette, Poythress, Norman G., Boothroyd, Roger A., Petrila, John and Mehra, Shabnam. (2005). Evaluating the efficiency and community safety goals of the Broward County Mental Health Court. *Behavioral Sciences and the Law, 23*: 227-243.
[5] Herinckx, Heidi A., Swart, Sandra C., Ama, Shane M., Dolezal, Cheri D. and King, Steve. (2005). Rearrest and linkage to mental health services among clients of the Clark County mental health court program. *Psychiatric Services, 56*(7): 853-857.
[6] O'Keefe, Kelly. (2006). The Brooklyn mental health court evaluation: Planning, implementation, courtroom dynamics, and participant outcomes. New York: Center for Court Innovation. Available online at: http://www.courtinnovation.org/_uploads/documents/BMHCevaluation.pdf. Last accessed 29 December 2007.
[7] Moore, Marlee E. and Hiday, Virginia A. (2006). Mental health court outcomes: A comparison of re-arrest and re-arrest severity between mental health court and traditional court participants. *Law and Human Behavior, 30*: 659-674.
[8] Van Vleet, Russell K., Hickert, Audrey O., Becker, Erin E. and Kunz, Chelsea. (2008). *Evaluation of the Salt Lake County Mental Health Court: Final Report*. Available online at: http://www.law.utah.edu/_studyfiles/94/94.pdf. Last accessed 16 November 2009.
[9] Ferguson, Andrew, McAuley, Kristen, Hornby, Helaine and Zeller, Dennis. (2008). *Outcomes from the Last Frontier: An Evaluation of the Palmer Coordinated Resources Project*. South Portland, ME: Hornby Zeller Associates, Inc. Available online at: http://www.mhtrust.org/documents_mhtrust/Palmer%20Mental%20Health%20Court%20Evaluation.pdf. Last accessed 16 November 2009.
[10] Christy *et al*, 2005.
[11] Neiswender, John R. (2004). Executive summary of evaluation of outcomes for King County mental health court. Seattle, WA: Washington State University. Available online at: http://www.metrokc.gov/KCDC/mhcsum32.pdf. Last accessed 29 December 2007.
[12] Trupin, Eric, and Richards, Henry. (2003). Seattle's mental health courts: Early indicators of effectiveness. *International Journal of Law and Psychiatry, 26*(1): 33-53.
[13] Frailing, Kelly. (In press). How mental health courts function: Outcomes and observations. *International Journal of Law and Psychiatry*.

ISBN 978-0-9566139-1-2

[14] McNiel, Dale E. and Binder, Renee L. (2007). Effectiveness of a mental health court in reducing criminal recidivism and violence. *American Journal of Psychiatry, 164*: 1395-1403.
[15] Hiday, Virginia A. and Ray, Bradley. (2010). Arrests two years after exiting a well- established mental health court. *Psychiatric Services, 61*(5): 1-6.
[16] Cosden, Merith, Ellens, Jeffrey, Schnell, Jeffrey, Yamini-Diouf, Yasmeen and Wolfe, Maren M. (2003). Evaluation of a mental health treatment court with assertive community treatment. *Behavioral Sciences and the Law, 21*: 415-427.
[17] Cosden, Merith, Ellens, Jeffrey, Schnell, Jeffrey, and Yamini-Diouf, Yasmeen. (2005). Efficacy of a mental health treatment court with assertive community treatment. *Behavioral Sciences and the Law, 23*: 199-214.
[18] Bess, Gary. (2004). Mentally ill offender crime reduction (MIOCR) grant: Butte County forensic research team (FOREST). Available online at: http://www.garybess.com/samples/Applied_Research_Mentally_Ill.pdf. Last accessed 29 December 2007.
[19] O'Keefe, 2006.
[20] Griffin, Patricia A., Steadman, Henry J. and Petrila, John. (2002). The use of criminal charges and sanctions in mental health courts. *Psychiatric Services, 53*(10): 1285-1289.
[21] Redlich, Allison D., Steadman, Henry J., Monahan, John, Petrila, John and Griffin, Patricia A. (2005). The second generation of mental health courts. *Psychology, Public Policy and Law, 11*(4): 527-538.
[22] Redlich, Allison D., Steadman, Henry J., Monahan, John, Robbins, Pamela C. and Petrila, John. (2006). Patterns of practice in mental health courts: A national survey. *Law and Human Behavior, 30*: 347-362.
[23] Poythress, Norman G., Petrila, John, McGaha, Annette and Boothroyd, Roger A. (2002). Perceived coercion and procedural justice in the Broward mental health court. *International Journal of Law and Psychiatry, 25*: 517-533.
[24] Stodola, Mark A. (2004). An assessment and evaluation of mental health courts in Maricopa County, Arizona. Williamsburg, VA: National Center for State Courts Institute for Court Management, Court Executive Development Phase III Paper.
[25] Bureau of Justice Assistance. (2007). Mental health court learning sites. Informational sheet on Washoe County, Nevada MHC. Available online at: http://consensusproject.org/downloads/Reno_snapshot.pdf. Last accessed 22 October 2007.
[26] Washoe County MHC Handbook. (2007). Handbook of the Washoe County Mental Health Court. Received via e-mail from Julie Vann on 1 November 2007.
[27] Seltzer, Tammy. (2005). Mental health courts: A misguided attempt to address the criminal justice system's unfair treatment of people with mental illness. *Psychology, Public Policy, and Law, 11*: 570-586.
[28] Parker, Lacey R. (2005). Mental health courts: Moving beyond the drug court model. *Developments in Mental Health Law, 24*(1): 1-15.
[29] Miller, Robert D. and Maier, Gary J. (1990). Factors affecting the decision to prosecute mental patients for criminal behavior. In D. Wexler (ed.). *Therapeutic Jurisprudence: The Law as a Therapeutic Agent*. Durham, NC: Carolina Academic Press: 369-377.
[30] Nevada Revised Statutes (NRS) 176A.260. (2001). Conditions and limitations on assignment of defendant to program; effect of violation of terms and conditions; discharge of defendant upon fulfillment of terms and conditions; effect of discharge.
[31] Albonetti, Celesta A. and Hepburn, John R. (1996). Prosecutorial discretion to defer criminalization: The effects of defendant's ascribed and achieved status characteristics. *Journal of Quantitative Criminology, 12*: 63-81.
[32] Albonetti, Celesta A. (1987). Prosecutorial discretion: The effects of uncertainty. *Law and Society Review, 21*: 291-313.
[33] Albonetti, Celesta A. (1986). Criminality, prosecutorial screening and uncertainty: Toward a theory of discretionary decision making in felony case proceedings. *Criminology, 24*: 623-644.
[34] Bonta, J., Law, M. and Hanson, K. (1998). The prediction of criminal and violent recidivism among mentally disordered offenders: A meta-analysis. *Psychological Bulletin, 123*: 123-142.
[35] Ulmer, Jeffery T. (2001). Intermediate sanctions: A comparative analysis of the probability and severity of recidivism. *Sociological Inquiry, 71*(2): 164-193.
[36] Lamb, H. R. and Weinberger, Linda E. (2008). Mental health courts as a way to provide treatment to violent persons with severe mental illness. *Journal of the American Medical Association, 300*(6): 722-724.
[37] Frailing, In press.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

A CRITICAL EVALUATION OF KEY ISSUES IN THE PHILOSOPHY OF SCIENCE THAT AFFECT THE METHODS AND THEORIES USED IN COGNITIVE NEUROSCIENCE

Helen GORDON

MSc Student, University of Nottingham, Nottinghamshire, UK
lwxhag@nottingham.ac.uk

Keywords: Epistemology, Metaphysics, Ethics, Neuro-imaging, Realism, Dualism, Qualia

ABSTRACT

Cognitive neuroscience is the scientific study of mental processes in the brain, and it is a particularly pertinent area of psychology to investigate the philosophical issues involved with the study and practice of it; because the results of experimentation in this field can inform philosophical thinking, for example: the relation between the mind and the body.[1] In this essay I will argue that the philosophical issues associated with theoretical and methodological concerns involved in studying cognitive neuroscience do not threaten its position as an empirical science. Furthermore, studying cognitive neuroscience can provide important and applicable knowledge about processes within the brain. I will emphasise the need to appreciate that cognitive neuroscience is a relatively new area of psychology, having only officially become an independent discipline in the 1970's[2] and is dependent upon new technology to gain its evidence, and therefore has potential for further development and verification to establish greater scientific rigour in the future.

INTRODUCTION

There are three main branches of philosophy that can be utilised to investigate the nature of psychology as a science; epistemology, metaphysics and ethics. I will critically evaluate each one and their impact upon the methods and theories of cognitive neuroscience.

Epistemic Concerns with Methods Used in Cognitive Neuroscience

Epistemology is concerned with how science can acquire any knowledge at all. There is a long and established history of philosophy of science; Popper[3] rejected classical empiricism and instead put forward the falsification criterion; a genuine science can be differentiated against a pseudo-science if its theories can be falsified. It is understood from the thinking of the logical positivists that scientific knowledge can be gained when researchers perform experiments that produce observable results that can be statistically summarised, and then theories can be proposed about the mental representations and processes that produce the different brain activities.[4]

However, one major concern in cognitive neuroscience is the problem of indirectness of inquiry; mental processes are inaccessible, any knowledge must be arrived at indirectly.[5] There are no direct forms of observation of brain processes, neuroscientists have to rely on techniques and instruments that could potentially generate artefacts and threaten the status of the knowledge we learn from these studies.

The methods used in neuroscience are lesion studies, cell recording and neuro-imaging, and I will discuss the limitations that each one has when allowing us to investigate the brain, before concluding that cautious use of such techniques can provide us with useful and practical knowledge about processes within the brain.

ISBN 978-0-9566139-1-2

Lesion studies are a widely used tool for relating brain structure to function by identifying the deficit associated with an area, and inferring that area's contribution to normal function. However, a lesioned brain is not equal to a normal brain minus the lesioned area, brains can develop alternative, compensatory pathways, as the brain is highly plastic.[6] Inferring normal function from a damaged system is highly problematic, the loss of cognitive function from an area following a lesion doesn't necessarily show what information processing was occurring in that area, thus the ability to construct an accurate theory about what function occurs in a specific area of the brain is threatened.

By relating lesion results to single cell recordings, deficits exposed by a lesion study can show what information is being processed on a cell recording, in an attempt to develop a diagram decomposing a system to its functional components for the brain.[7] Cell recording involves inserting an electrode into an individual neuron and correlating their firing rate with features of the tasks they are responding to, so that if a cell is particularly active in response to a certain stimuli, it suggests the cell processes information about the stimuli.

However, there are problems regarding this method: firstly, brains are made up of thousands of neurons of different types, any conclusions drawn are limited to only stimulus-response correlations.[8] It requires further inference to conclude that the cell is processing that information rather than simply conducting it, or feeding it back to another brain area.[9] Secondly, this method will pick up accidental voltages from other cells; it is too imprecise to directly connect individual cell recording with larger brain activation patterns, which is the aim of experimentation and theory-making in cognitive neuroscience. Thirdly, cell recording only focuses on one very particular area of the brain at one time, missing out on potential processes occurring simultaneously in other parts of the brain.

Neuro-imaging can record from multiple areas of the brain at once, thus overcoming the latter issue mentioned for cell recording.[10] Neuro-imaging such as fMRI (functional magnetic resonance imaging) scans measure increased blood flow in an area of the brain while performing certain tasks and reveal these areas with coloured blocks on a pictorial representation. The definiteness of the images is a positive of this approach, the 'lit up' areas have been suggested to be robust for certain tasks, and their findings are consistent with lesion studies and cell recordings.[11]

However, there is no way of determining whether the differences found are due to the cognitive processes or another concurrently occurring event[12], causing theoretical issues when it comes to theory-making and ascribing a cause to an effect. Furthermore, adding to this issue, there is no normative standard of a brain measure, overall size and size of specific regions are merely suggested. A brain region is deemed to be 'active' when a statistical threshold of certainty is crossed, an arbitrary level designed to balance Type I and Type II errors[13], which is projected over a brain template that may not even belong to the individual being scanned.[14]

The basic assumptions of these neuroimaging and cell recording studies, that they assume localisation of brain function, neurons are the basic functional unit of the brain and different brain areas perform different processing operations[15], illustrate the modularity of mind hypothesis.[16] This states that the mind is a system of specialised modules designed for specific tasks and this information is encapsulated, immune from information from other sources. However, this theory is controversial, and as mentioned above, the brain has multiple firing patterns and constant background activity that keeps the modularity of mind hypothesis from being empirically verified.[17]

Furthermore, it must be noted that there is a theoretical gap between raw data recordings and interpretable data, and Thagard[18] comments '*scientists use what they know to cull data that supports what they believe to be the case*' (p. 309), that is, a theory can guide interpretation of the imaging results: theory-dependant observation. This seems to go against the traditional empirical view of science where data is the foundation of developing theories. However, Bechtel[19] maintains that this practice is common in neuroscience, and can prevent the suspicion of results being due to artefacts. If there was a failure to find a theory based on the results of investigation, scientists assume that the techniques used are generating artefacts, but if a theory can be found to understand the results of such techniques, then it can oppose the danger of artefact generation in such techniques[20], thus supporting a verificationist instead of a falsificationist view of science, a view originally supported by the logical positivists and advanced by Ayer.[21]

ISBN 978-0-9566139-1-2

Despite these difficulties, I do not believe they threaten cognitive neuroscience's status as a science; neuroimaging can provide us with reliable knowledge about processes within the brain, and this can have practical applications for treating malfunctions in brain processes, such as epilepsy. Studies investigating the effectiveness of neuroimaging techniques used to visualise epileptic seizure focus prior to surgery show that 70-100% of scans were correctly localised in patients with temporal lobe epilepsy.[22] Also, patients with a correctly localised scan had a significantly better outcome following surgery than those with a non-localised scan.[23] Without theories of localisation of brain function and modularity of mind, surgery for sufferers of epilepsy would be severely hindered by inaccuracy. Though the techniques used to gain knowledge about the brain may have several drawbacks, we are able to gain enough knowledge that is proved to be accurate when we apply the theories to such real-life situations, the theories are referentially successful in that they allow a positive contribution to improving a person's health and quality of life to be made. Indeed, Bechtel[24] maintains that all three procedures should be used and aligned to produce converging results, thus showing that the process under investigation is not an artefact. He argues that it is unlikely that, by chance, two different techniques would produce the same results, allowing verification of other findings and theories.

This utilitarian approach emphasises the pragmatic advantages in using the methods above to gain knowledge about mental processes. It is important to note that neuroimaging techniques, such as fMRI have only being refined in the 1990's.[25] Much future development of neuroimaging techniques may be able to close the debate around the difficulties generated by its current theoretical and methodological issues, but for the present, a pragmatic, patient approach to the converging results from investigations into the brain can be used to gain genuine scientific knowledge.

Metaphysical Issues of the Theoretical Assumptions of Cognitive Neuroscience

Modularity of mind and brain localisation are basic theoretical assumptions of cognitive neuroscience. However, there is a question surrounding what constitutes a module, and where the boundaries of certain modules lie. This theory is fraught with difficulties that introduce the metaphysical issues involved in cognitive neuroscience, concerning the existence and the nature of different kinds of entities[26], and whether we are justified in believing in the existence of theoretical objects that are hypothesised and not directly observed.

The first metaphysical issue is that of scientific realism and the nature of unobservables: mental processes that cannot be directly observed by any means, including neuroimaging. The position of realism maintains for a neuroscientific theory to be true, it is a necessary condition that the entities in the theory do exist[27], sciences aims to provide a true description of the world, and scientific knowledge is not limited by our powers of observation. Following this, unobservables are the same as observable entities, they are genuinely existing objects; the highlighted areas on a fMRI scan are true representations of genuinely existing mental processes. I will not discuss the issue of representation here, but there has been debate over the nature of the relation of representation, their embodiment in the brain and how mental operations involve the manipulation of internally constructed representations of external phenomena.[28]

The anti-realist position opposes the realist view about unobservable entities and states that knowledge of the unobservable part of the world cannot be gained, theories are not literal descriptions of reality, but idealisations that bare no correlation to the real world. An anti-realist would view theories about unobserved objects as being underdetermined by data; anything we have learnt from neuroimaging studies and cell recording could not be seen as true scientific knowledge. A realist would respond to this by claiming that underdetermination is no barrier for acquiring knowledge; it does not men science cannot give us knowledge of unobservables, even though we cannot directly see what is occurring in the brain, we can successfully use findings from neuroimaging to make predictions that can be verified.[29]

The 'No Miracles Argument' can soundly undermine the anti-realist position; our successful theories about neuroscience make predictions that would seem miraculous unless it is assumed that they do refer to objectively existing processes and are approximately true[30], like the surgery on sufferers of temporal lobe epilepsy discussed above[31]. Therefore, I maintain that realism should be accepted, as it sufficiently explains the relative success of such techniques, and reject the anti-realist threat to cognitive neuroscience's scientific standing.[32]

ISBN 978-0-9566139-1-2

The second metaphysical issue is that of the renowned and ubiquitous mind/body problem; the relation of the mind, and the mental world, to the brain and the physical world. There are two key ways to make sense of the problem. Dualism maintains that the mental and the physical are two separate, fundamentally different entities, that the mind is distinct from the physical body and that the mental dimension cannot be reduced to the physical.[33] This view can seem initially incompatible with the basic tenets of cognitive neuroscience, as this studies the biological foundations of mental phenomena, how the mind is instantiated in the brain and how structure and processes in the brain enable cognitive processes in the mind.[34] This seems to answer the mind/body debate in reductionist terms, presenting a materialist viewpoint; that mental events are nothing more than physical events in the brain, mental states are identical to brain states;[35] cognitive neuroscience purports a materialist viewpoint.

However, there is a gap between explanations of how the brain enables information processing of the mind, but is unable to explain the qualitative, subjective experiences that come with it, or the theory of consciousness.[36] For example, the theory of visual processing of colour cannot explain the sensory experience of 'red'. These subjective qualities of conscious experiences are known as qualia, and a materialist would have to deny the existence of qualia as something that hinders access to knowledge and supports dualism; the implication would be that one would have to go beyond the study of the structure and processes in the brain to fully understand human behaviour, and that some other substance or property exists that has influence over our behaviour.

I maintain that cognitive neuroscience should not be used as proof of, or in equivalence to, the materialist paradigm; there has not been found a neural correlate of consciousness yet, so all possible explanations of this phenomena should be investigated. Our subjective state of mind can influence how effectively we can recover from diseases[37], a non-material force can change the wiring of the brain through focused attention or directed mental effort, a concept known as self-directed neuroplasticity.[38] Hence, the interaction between theories of neuroscience and theories of philosophy of science can be illustrated. The materialist paradigm fails to give an adequate account for all mechanisms operating within the brain, and thus presents a theoretical restriction, further study is needed in this controversial area, before any impact upon the scientific standing of cognitive neuroscience should be debated.

Ethical Issues Involved with the Application of Cognitive Neuroscience

Ethics is concerned with how people ought to behave; a normative rather than descriptive branch of philosophy. Ethics is not replaceable by cognitive science, nor is it reducible to it.[39] There are ethical issues involved with doing psychological research, as the subject of investigation is a person, and basic principles set down by the British Psycholigcal Society (BPS) must be adhered to when conducting any form of experiment. Informed consent must be obtained, deception should be avoided, and respect of the individual's confidentially and well-being must be at the forefront of experimenters mind at all times.

Neuroimaging assumes that psychological traits have physical correlates that are measurable with brain imaging technology.[40] This relationship is illustrated in unconscious racial attitudes; prejudice and out-group exclusion can be addressed in terms of underlying brain function. Phelps et al.[41] found disassociation between implicit and explicit racial bias; white people, shown pictures of unfamiliar black faces, displayed an increase in amygdala activation, which was related to unconscious social evaluation, but as this race-bias response is totally separate from consciously held attitudes, the extent to which a person is responsible for socially unacceptable views and behaviour becomes unclear.

There are also broader issues associated with moral blame and implications for what we consider to be right and wrong when we utilise the findings of research in neuroscience. Morality assumes someone is in the wrong and a failure of their character is to blame for their omission. Neuroimaging can show the biological cause in the brain of this failure; cognitive neuroscience could show why people are amoral.

Neuroimaging can also be utilised in diagnosing affective disorders; in predicting violent, antisocial behaviour before it surfaces. Anti-social Personality Disorder (APD), as defined by Axis II of the Diagnostics and Statistical Manual IV Revised (DSM-IV R), is diagnosed by the inability to process emotional signals and feel emotions such as guilt and empathy. Dysfunction of the amygdala is associated with aggression and explains why sufferers of APD have diminished responsiveness to facial expressions.[42] This brings up issues regarding responsibility

ISBN 978-0-9566139-1-2

and criminal culpability. Humans are perceived to be intentional, rational agents, such findings fuel deterministic accounts of human behaviour in the ubiquitous free will/determinism debate; an individual may not be responsible for any crimes committed if they have psychopathic, aggressive tendencies 'built' into their brain. Illes and Raffin[43] prescribe caution against the dangers of 'runaway' use and premature application outside of research settings that could lead to unethical use, and inflict risks on society. Canli and Amin[44] also warn against naïve acceptance of brain scans as proof of an individual's brain state.

Nevertheless, neuroimaging has many benefits if used with care; it allows for the identification of vulnerability factors for psychopathology and diagnosis of disorders, identifying those who could benefit from preventative treatment, such a cognitive-behavioural therapy, identification of pathological brain areas that need surgery, and therefore has a positive role as it can additively contribute towards people's well-being and mental health.

CONCLUSION

The methods used in cognitive neuroscience such as neuroimaging come up against many philosophical issues, such as the indirect nature of investigation, concerns about brain localisation theory, and the nature of the existence of mentally occurring phenomena. However, as I have argued throughout this essay, cognitive neuroscience is still a legitimate, empirical science, it is evidence-based, and this evidence refers successfully to processes in the brain. Furthermore, the study of cognitive neuroscience can help to understand the brain on a greater level, and can serve practical, positive uses when providing treatment or care for disorders of the brain.

There is a need to monitor how this new knowledge is applied, to avoid unethical use, and certain topics need to be researched in greater depth to discover more robust findings about the processes of the brain. However, the study of brain functions is the study of the psychological processes of organisms, and the imaging and recording gains meaning by being related to a theory of what is occurring functionally.[45] There is a positive value in critically evaluating the scientific claims that brain function theories are based, but it is unwarranted to falsely promote scepticism in a relatively new and dynamic area of psychology that is still developing and refining it's investigative techniques.

REFERENCES

[1] Bechtel, W. *From imaging to believing: Epistemic issues in generating biological data.* (2000). In R Creath & J Maienschein (eds.) Epistemology & Biology, Cambridge University Press, pp. 138-63.
2 Gazzaniga, M. S., Ivry, R. B. & Mangun, G. R. *Cognitive Neuroscience: The biology of the mind.* (2002). New York: W.W.Norton.
[3] Popper, K. *The Logic of Discovery.* (1959). London: Hutchinson.
4 Thagard, P. *Philosophy of Psychology and Cognitive Science.* (2007). North Holland Elsevier.
[5] Bechtel, W., Mandik, P., Mundale, J. & Stufflebeam, R. *Philosophy & the Neurosciences: A Reader.* (2001) Blackwell Publishing.
[6] Merzenich, M., Kaas, J., Sur, M., Nelson, R., & Felleman, D. Progression of change following median nerve section in the cortical representation of the hand areas 3b and 1 in an adult owl and squirrel monkeys. (1983). *Neuroscience*, 10(3) pp. 639-665.
[7] Thagard, P. *Philosophy of Psychology and Cognitive Science.* (2007). North Holland Elsevier.
[8] Thagard, P. Philosophy of Psychology and Cognitive Science. (2007). North Holland Elsevier.
[9] Bechtel, W. Aligning Multiple Research Techniques in Cognitive Neuroscience: Why Is It Important? (2002). *Philosophy of Psychology*, 69, S48-S58.
[10] Bechtel, W. Aligning Multiple Research Techniques in Cognitive Neuroscience: Why Is It Important? (2002). *Philosophy of Psychology*, 69, S48-S58.
[11] Bechtel, W., Mandik, P., Mundale, J. & Stufflebeam, R. *Philosophy & the Neurosciences: A Reader.* (2001) Blackwell Publishing.
[12] Thagard, P. *Philosophy of Psychology and Cognitive Science.* (2007). North Holland Elsevier.
[13] Nicholas, B. *Neuroethics: A Literature Review prepared for the Toi te Taiao: the Bioethics Committee.* (2006). New Zealand.
[14] Wolpe, P., Foster, K., Langleben, D. Emerging Neurotechnologies for Lie Detection: Promises and Perils. (2005). *The American Journal of Bioethics.* 5(2), pp. 39-49.
[15] Bechtel, W & Richardson, R. *Discovering Complexity: Decomposition and Localisation as Scientific Research Strategies.* (1993). US: Princeton University Press.
[16] Foder, J. *Modularity of Mind: An Essay on Faculty Psychology.* (1983). MIT Press.
[17] Thagard, P. *Philosophy of Psychology and Cognitive Science.* (2007). North Holland Elsevier.

ISBN 978-0-9566139-1-2

[18] Thagard, P. *Philosophy of Psychology and Cognitive Science.* (2007). North Holland Elsevier.
[19] Bechtel, W. *From imaging to believing: Epistemic issues in generating biological data.* (2000). In R Creath & J Maienschein (eds.) Epistemology & Biology, Cambridge University Press, pp. 138-63.
[20] Bechtel, W., Mandik, P., Mundale, J. & Stufflebeam, R. *Philosophy & the Neurosciences: A Reader.* (2001) Blackwell Publishing.
[21] Ayer, A. *Language, Truth & Logic.* (1936). New York: Oxford University Press.
[22] Whiting, P., Gupta, R., Burch, J., Mujica Mota, R.E., Wright, K., Marson, A., Weishmann, U., Haycox, A., Kleijnen, J. and Forbes, C. A systematic review of the effectiveness and cost-effectiveness of neuroimaging assessments used to visualise the seizure focus in people with refractory epilepsy being considered for surgery. (2006). *Health Technology Assessment* 10(4).
[23] Whiting, P., Gupta, R., Burch, J., Mujica Mota, R.E., Wright, K., Marson, A., Weishmann, U., Haycox, A., Kleijnen, J. and Forbes, C. A systematic review of the effectiveness and cost-effectiveness of neuroimaging assessments used to visualise the seizure focus in people with refractory epilepsy being considered for surgery. (2006). *Health Technology Assessment* 10(4).
[24] Bechtel, W. Aligning Multiple Research Techniques in Cognitive Neuroscience: Why Is It Important? (2002). *Philosophy of Psychology*, 69, S48-S58.
[25] Gazzaniga, M. S., Ivry, R. B. & Mangun, G. R. *Cognitive Neuroscience: The biology of the mind.* (2002).New York: W.W.Norton.
[26] Okasha, S. *Philosophy of Science: A Very Short Introduction.* (2002). Oxford Paperbacks.
[27] Okasha, S. *Philosophy of Science: A Very Short Introduction.* (2002). Oxford Paperbacks.
[28] Churchland, P. *Neurophilosophy: towards a Unified Science of the Mind-Brain.* (1988). MIT Press.
[29] Okasha, S. *Philosophy of Science: A Very Short Introduction.* (2002). Oxford Paperbacks.
[30] Putnam, H. What is mathematical truth? In *Mathematics, Matter and Method. Philosophical Papers*, 1. (1975). Cambridge University Press, pp 60-78.
[31] Whiting, P., Gupta, R., Burch, J., Mujica Mota, R.E., Wright, K., Marson, A., Weishmann, U., Haycox, A., Kleijnen, J. and Forbes, C. A systematic review of the effectiveness and cost-effectiveness of neuroimaging assessments used to visualise the seizure focus in people with refractory epilepsy being considered for surgery. (2006). *Health Technology Assessment* 10(4).
[32] Okasha, S. *Philosophy of Science: A Very Short Introduction.* (2002). Oxford Paperbacks.
[33] Churchland, P. *Neurophilosophy: towards a Unified Science of the Mind-Brain.* (1988). MIT Press.
[34] Repovš, G. Cognitive Neuroscience and the Mind/body Problem. (2002). *Horizons of Psychology*, 13(2) pp. 9-16.
[35] Bechtel, W., Mandik, P., Mundale, J. & Stufflebeam, R. *Philosophy & the Neurosciences: A Reader.* (2001) Blackwell Publishing.
[36] Chalmers, D. Facing up to the problem of consciousness. (1995). *Journal of consciousness studies*, 2(3) pp. 200-219.
[37] Sternberg, E. & Gold, P. The Mind-Body Interaction in Disease. (2002). *Scientific American Special Edition*, 12(1) pp: 82-129.
[38] Schwartz, J., Stapp., H. & Beauregard, M. Quantum physics in neuroscience and psychology: a neurophysical model of mind–brain interaction. (2004). *Philosophical Transactions of the Royal Society*, B, pp. 1-19.
[39] Thagard, P. *Philosophy of Psychology and Cognitive Science.* (2007). North Holland Elsevier.
[40] Farah, M. Neuroethics: the practical and the philosophical. (2005). *Trends in Cognitive Sciences*, 9(1), pp. 34-40.
[41] Phelps, E.A., O'Connor, K., Cunningham., W., Funayama, E., Gatenby, J., Gore, J., and Banaji, M. Performance on indirect measures of race evaluation predicts amygdala activation. (2000). *Journal of Cognitive Neuroscience*, 12(5), pp. 729-738.
[42] Blair, R. Neurocognitive models of aggression, the anti-social personality disorders and psychopathy. (2001). *Journal of Neurology, Neurosurgery and Psychiatry*. 71(6), PP. 727-731.
[43] Illes, J., and Raffin, T. Neuroethics: An emerging new discipline in the study of brain and cognition. (2002). *Brain and Cognition*. 50(3), pp. 341-344.
[44] Canli, T. & Amin, Z. Neuroimaging of Emotion and Personality: Scientific Evidence and Ethical Considerations. (2002). *Brain and Cognition*. 50(3), pp. 414-431.
[45] Hatfield, G. *Perception and Cognition: Essays in the Philosophy of Psychology.* (2009). Oxford University Press.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

APOLLO'S AMOURS: TILL DEATH DO US PART? HOMOEROTIC MYTHOLOGY IN THE ITALIAN RENAISSANCE: A Cultural and Psychoanalytical Inquiry

Ann HAUGHTON

Post-Graduate Research Student, History of Art, University of Bristol, UK.
ah5415@bris.ac.uk

Such indeed is the importance of the subject that it still calls for fresh investigation, and may be studied with advantage from most points of view.
Jacob Burckhardt: The Civilisation of the Renaissance in Italy (1860)

ABSTRACT

Recent decades have witnessed an abundance of scholarly and interpretive perspectives brought to bear upon lesbian and gay history, culture and identity in the field of modern and contemporary visual representation. However, homoerotic depiction of social and sexual activities between men also permeates the visual record of the Italian Renaissance. This paper will explore the integral role of specific visual representations of homoerotic myths which conflate erotic desire and philosophical allegory, namely Benvenuto Cellini's marble Apollo and Hyacinth (1545) and Giulio Romano's drawing of Apollo and Cyparissus (1524) as records and proclamations of contemporary power relations related to the constructs and dynamics of gender, identity and sexuality. Both these images were executed at a time when erotic relationships between men found conscious expression in the narrative and imagery of pagan mythology as an exclusive model sanctified by antique precedent. But although both display an undeniable erotic component, these visual texts reveal far more than lascivious experiences shared by males in Renaissance Italy. There are also pedagogical, initiatory and moral values encoded within these works as a means of training and socialization of Renaissance adolescents. As images which embody recognized cultural parameters and norms relating to gender roles and sexual desire that differ widely from our own modern notions of social equality or reciprocity, they are a rich source for understanding the social, political, institutional and cultural contexts that underpinned their production and reception.

This paper aims to elucidate new and rewarding opportunities for addressing the issue of homoeroticism in the Italian Renaissance through dedicated historically informed enquiry. It will explore the ways which art of this nature challenges notions that heterosexuality was normative and universal in the early-modern period with consideration of the taxonomical or culturally specific values that underpinned its execution. The understanding and contextualizing of the dynamics of same-sex eroticism and sociality, as represented in the artistic production of the cinquecento is an important but so far relatively neglected area of study within the field of art historical discourse and it is hoped that further research will help to edge this fascinating subject closer to the intellectual mainstream.

INTRODUCTION

Personal sexual behaviour is shaped by and shapes the wider social and political milieu, but not all societies permit expression of all varieties of erotic disposition. Men who characteristically prefer relations with youths are considered in our culture deserving of sanction, if not outright condemnation, but in many other cultures,

ISBN 978-0-9566139-1-2

particularly classical antiquity and the Italian Renaissance, age-asymmetrical relations were considered to be a transient and natural stage in the lives of both adults and youths. For these cultures adolescence was treated as a subset of adulthood, not signalling full maturity but nevertheless sharply distinguishable from childhood. Judicial records testify that at one time or another and with varying significance and degrees of involvement, pederastic relations formed part of the life experience of many Italian males of the late medieval and early modern period. Rediscovered artefacts and textual sources such as Ovid's poem Metamorphoses (8AD),[1] which provides the mythic narrative for the images discussed in this paper, revealed to the Italian upper classes a culture which faced directly and candidly a fundamental aspect of human experience - men sometimes love, whether spiritually or sexually, adolescent males. However, this was also an age when civil morality and theological condemnation of male sexual relations prompted social, moral and political divergences and invectives over such activity.[2] As a direct consequence of this condemnation, erotic relationships between men found conscious expression in the narrative and imagery of pagan mythology as an exclusive model sanctified by antique precedent. This paper will present new interpretations of homoerotic mythological representations of Apollo with his younger paramours Hyacinth and Cyparissus as structured initiatory and pedagogical models connected to rites that mark the passage from youth to adulthood. It will also address the instrumentality of these images in terms of understanding the social articulation of Renaissance power differentials and constructs, as well as issues of sexuality, identity and gender.

DISCUSSION

The demarcation should be clearly and accurately drawn between the pederasty that existed in the Renaissance and that of the paedophilia that pervades our modern world. At the time these images were executed pederastic relations were confined to pubescent boys, preferably between the ages of twelve and eighteen. There is no evidence to suggest that pre-pubescent boys attracted desire at all. In both antiquity and Renaissance Italy such behaviour would have been roundly condemned as reprehensible in that it served no pedagogical or formative purpose. Therefore, it can be assumed that age differentials were often idealised in homoerotic imagery since in reality when it came to pederastic bonds there was less disparity in age than in marriages of the period when the girl was of an average age of thirteen and the suitor around thirty.

Apollo is prominently placed in Ovid's Metamorphoses as one of the most important Olympian gods, who as the eternal beardless divinity had the most prominent and prolific male relationships of all, but the two romances which feature most commonly in Renaissance artistic production are with Hyacinthus and Cyparissus, both of whom died at his hands. Cellini's marble sculpture Apollo and Hyacinth[3] features the divinity with his beloved mortal youth Hyacinth who in Ovid's text when being taught how to throw the discus by his mentor and lover was struck as it fell to the ground and died. Apollo refused to allow Hades to claim the young man; rather, he made a flower, the hyacinth, from his spilled blood. In his pen and ink drawing of Apollo and Cyparissus,[4] Romano places the deity with his other beloved young mortal Cyparissus. The drawing was engraved by Marcantonio Raimondi and distributed in printed form, but these are no longer extant. Apollo adored Cyparissus dearly also and bequeathed to his juvenile beloved a beautiful tame stag. When Cyparissus accidentally killed his stag whilst being educated in the manly art of hunting he was distraught by his loss. All of Apollo's consolations were in vain and Cyparissus was so distressed that he begged to be allowed to mourn forever. Eventually Apollo obliged by turning him into a cypress tree which is said to be a sad tree because of the droplets of sap that form on its trunk.

There were different iconic ambits in the cinquecento - one private the other public, and homoerotically charged imagery assumed different faces for private and public consumption. Cellini's statue was sculptured in anticipation of its acceptance by Cosimo de Medici who donated the marble but ultimately refused it after the artist was imprisoned for his own sexual transgressions. Romano executed his drawing when planning frescoes for Baldassarre Turini but the commission was unfulfilled and the drawing was subsequently issued as prints - all of which have been destroyed. Although these images were executed in very different media and for diverse viewing conditions, they are superb exemplars of how Renaissance mythological representations with unmistakably homoerotic sentiments contained a complex set of encoded messages that capture the physical, emotional and social elements of intergenerational same-sex male experience. The interconnecting and over arching theme present in both images is their strong pedagogical and initiatory function correlating rites that mark the passage from youth to adulthood. There has been a paucity of scholarly discourse on these two images

ISBN 978-0-9566139-1-2

to date, and the little that does exist barely acknowledges the homoerotic narrative which inspired these works. In the case of Cellini's sculpture, Michael Cole dismisses the subject as an '*infelicitous, ad hoc, and even nonsensical idea adopted in the face of the difficulties the marble presented*' with '*no narrative or thematic rationale to Hyacinth's presence*'.[5] And in her seminal work on Romano's erotic prints Bette Talvacchia does not make adequate distinction between his homoerotic subject matter and his rendering of heterosexual couplings.[6] Any attempt to coyly 'straighten-out' these works with heterocentric exclusion of their prevailing homoerotic characteristics fails to do justice to the spirit of their execution and to the mythological narrative from which they are sourced. Therefore, an alternative interpretation to the more conventional heteronormative scholarship currently published will be offered; one which has particular focus upon the rigid behavioural expectations encoded within these Renaissance depictions of male love, but which adopts a more appropriate perspective drawing upon queer theoretical analysis - to date the almost exclusive domain of literary and art historical discourse concerned with work produced in recent decades.

Closer reading of Apollo and Hyacinth reveals that Cellini captures the moment when Hyacinth is accidentally slain by the lethal discus that Apollo still holds in his left hand, preserving in sculptured form the precise moment when the juvenile Hyacinth is fatally injured but yet to transform into his new adult life form. Executed at a time when lived eroticism conformed to rules of social hierarchy with sexual roles tied to age as well as class, Cellini's Apollo and Hyacinth provided a visual allegory with which Renaissance men could have identified, presenting a precedent for sexualised relationships between males with the inference that such experiences were consistent with similar rigid codes of behaviour that dominated male societies in classical antiquity. Cellini reinforces this claim to classical tradition by seemingly taking as his point of reference the stance, hairstyle and physique of Apollo Belvedere, (c.130-140A.D)[7] as an established model of virtue triumphant in monumental figural form. His all antica Apollo captures the finest qualities of an honourable and powerful Renaissance patriarchal polity where strength, courage, vitality, nobility, energy and intelligence were expected behavioural codes. Sexual norms were, however, dependant on social status, gender, wealth and age in both classical antiquity and Renaissance Italy, thus Cellini's Apollo and Hyacinth presents a strategy of vertical hierarchy that reinforces Apollo's domination of the acquiescent Hyacinth by positioning the juvenile behind rather than in front or even at the side of his master and mentor. However, he does so in a manner that eschews corporeal strength in favour of a more intimate and implicit sense of subjugation which demarcates appropriate gender and power constructs but also encapsulates the poignancy of their parting. In virile hegemonic Florence the imposition of one's will would have been considered a defining characteristic. Sexual ethics and behaviour were governed not by the hetero/homosexual context but by the question of active / passive roles that were enmeshed with important comportmental codes. In contrast to the developed physique of the contrapposto Apollo staring straight ahead whilst proudly displaying the attributes of manhood, Cellini presents a kneeling pubescent Hyacinth, almost rooted to the ground whilst gazing adoringly up as if awaiting favour to be granted rather than expecting sexual delectation. For contemporary viewers, this would have transmitted far more than a suggestive representation of titillation and scurrilous erotic innuendo – it would have visually validated the edifying importance of patriarchal order and prescriptive ideals.

In terms of masculinity, as long as the adult protagonist took the sexually dominant role in his relations with boys, his sexuality would have done nothing to distract from his perceived maleness. To be identified as the older, active penetrator of an adolescent boy did not tarnish masculine identities; on the contrary it affirmed manliness, virility and honour. As a visual model which closely followed the dynamics and constructs of idealised male relationships in antiquity, Apollo like all deities in ancient mythology is depicted as the active partner who takes the initiative and obtains sexual gratification. And for Cellini himself, this personally significant work validated to his critics that a love directed at members of one's own sex is true of the gods as well. In short, dominant contemporary definitions of masculinity remain preserved in Cellini's Apollo and Hyacinth despite its homoerotic sentiment. It convenes with phallic confirmation of the sociopolitical supremacy of adult citizen males, with each partner taking, expected to take, and wishing to be perceived as taking a prescribed role.

James Saslow sees Cellini's arrangement of his figures as '*not overtly sexual with Apollo playing with the boy's hair*'.[8] However, in order to achieve a comprehensive reading of this work and the context of its execution, it would be erroneous to ignore the import of its powerful visual reference to carnal intent. There is little point in deconstructing this image without recognition of Cellini's allusion to the act of fellatio when he positions the younger protagonist's head on the same plane as Apollo's genitalia and guided firmly by the master's hand as he

ISBN 978-0-9566139-1-2

acquiescently kneels in supplication with sensually parted lips. However, true to his notoriously rebellious nature Cellini has subverted contemporary practices specific to this act, because the judicial records testify that a pattern existed as the norm for fellatio where it was the juvenile's penis that was inserted in the older agent's mouth. Unlike anal penetration which was deemed effeminizing, an adult partner who allowed a juvenile protégé to penetrate his mouth would not have been perceived as violating the gender order or as subverting prescribed phallic-centered conceptions of behavioural erotics because it is he who is taking charge of the act and its outcome. The recipient of semen in anal intercourse is passive but in fellatio he is active. Thus the act was not deemed one of supplication because the deed of taking the youth in his mouth would make the dominant older man the active agent in control and the one who can choose to give pleasure and take it away at will. The inclusion of fellatio in this work further underscores its reproductive allegory because from the experiences passed down from the active agent to his protégé - a new life of adulthood is born. In effect their union is a reproductive one of two males creating one being; one which has propagated spiritual offspring such as virtue, experience and knowledge - products with greater longevity than biological offspring. With his transgressive variation on the act of fellatio, the question of whether Cellini is articulating his own erotic predilections or blurring the active/passive division that existed in the case of fellatio in order to render it more congruent with heterosexual practice to avoid censure is open to conjecture. In his personal life Cellini was imprisoned because he had a longstanding, co-habiting relationship with another adult male. So perhaps Cellini wished to demonstrate that as the dominant male, Apollo, like himself, can exercise his prerogative concerning erotic practices and dictate whatever form the act of fellatio might take.[9] Nevertheless, there is little ambiguity about how the artist has didactically encapsulated social articulation of Renaissance power dynamics and constructs, as well as the important issues of sexuality, identity and gender that prevailed at this time in this work.

By contrast, the physical nature of Apollo's relationship with Cyparissus is unequivocal in Romano's drawing. However, this paper will consider a reading that looks beyond what might be perceived as inappropriate licentiousness by our modern tastes. Romano draws our attention to the youth's penis with Apollo's hand-to-genital gesture which might be perceived as erotic phallic pleasuring. But as already discussed these works can be read as allegories for the ending of adolescence and entry into the lived experience of procreative manhood where the imminent death of Apollo's young lovers are understood as mystical, symbolic and initiatory. In which case, Romano's pedagogic and formative Apollo could in fact be gesturing to his initiate's developing penis in a manner that indicates how Cyparissus as a young male is himself budding and maturing into an active agent with sexual capabilities of his own.

The touching, protection or presentation of male genitalia is far from the exclusive province of representations of pagan subject matter. Leo Steinberg convincingly asserts in his seminal work *The Sexuality of Christ in Renaissance Art and in Modern Oblivion* that Renaissance artists made Christ's genitalia the focal point of their religious depictions because it emphasized his sexuality which in turn symbolized his humanity.[10] Steinberg interprets this gesturing in the context of devotion to the Holy foreskin and the Feast of Circumcision which were important tenets of Christian faith fundamental to understanding the mystery of the Incarnation and its identification with humankind's redemption. Therefore, it could perhaps be argued that Romano's profane image of Apollo gesturing to his beloved's penis is less lewd in intent than might first appear. Apollo's gesture finds correlation in Mary's similarly demonstrative emphasis of the Christchild's genitals as depicted in Cariani's Madonna and Child with Donor, 1520, and Domenico Ghirlandaio's Adoration of the Magi, 1497.[11] Ultimately, whether the subject matter is sacred or profane the gesture of presenting the younger agent's penis can be read as signifying the creation of a sexually defined man. Mary presents Christ's genitalia as evidence of his manhood and Apollo could similarly be indicating the juvenile's own impending manhood in the same manner.

Comparison of Romano's classicised homoeroticism with his notoriously explicit I Modi reveals a greater level of decorum for even the implication of libidinous acts between men than those which portray actual penetrative sex between a man and a woman.[12] Romano is far more circumspect in Apollo and Cyparissus and couches their relationship with greater mythological allusion - even though the action is not nearly so lewdly explicit. This is particularly surprising in light of how printed matter offered a forum for explicit erotica to be discretely disseminated to an expanded audience defined by income, erudition and the privileges of status – the same spectatorship that was, according to legislative statistics, the most likely to be active practitioners of pederastic relationships. Apollo as the older partner is the insertive agent but Romano has no need to expose the actual act of phallic penetration because the implication of its presence alone is sufficient to reiterate its function as an

ISBN 978-0-9566139-1-2

index of hierarchical sociopolitical empowerment. In order to imply, rather than lewdly depict sexual congress, drapery separates the two protagonists as if to spare Cyparissus the effeminising humiliation of actual penetration. Thus his future status as a mature citizen ready to take his place in society remains preserved from compromise. Whereas he displays no such reticence in his almost perfunctory handling of heterosexual penetrative couplings, Romano instead cautiously deploys a 'slung-leg' motif to symbolise rather than depict sexual congress.

The erotic tenor of the slung leg as a potent esoteric sexual metaphor became increasingly recognised in Italy during the sixteenth century as a token for marital union, male appropriation and sexual possessiveness or compliance. However, this rhetoric of carnal gesture was not the exclusive province of mythological representations. As Romano's own Madonna and Child (1522-3)[13] demonstrates the period witnessed the emergence in popularity of this metaphoric slung leg idiom in association with devotional art. The Christchild bestriding the Virgin's thigh, in this instance, has no perverse connotation since it can be read as the consummation of Christ the bridegroom to Mary as the Church in accordance with the writings of St. Gregory the Great.

As discussed, we should view any possible congruence between Romano's Apollo and Cyparissus and the lewd libertinism of his I Modi as tenuously held together by an erotic thread. The 'slung-leg' was used by Raphael in a biblical context in 1519 for his fresco on the ceiling vault of the Palazzi Pontifici, and there is greater consonance between Romano's drawing and his Old Testament scene of Isaac and Rebecca sharing an intimate embrace.[14] Raphael employs this symbolic device to portray conjugal union and also includes a third party voyeur. When we consider that Romano is recorded as having worked on these frescoes while apprenticed to Raphael, it is quite plausible that he may have 'queered' his master's heteronormative and biblically themed composition executed on the ceiling vault of the papal apartments for his own homoerotically charged mythological design at his socially ambitious friend's villa. Just as Raphael 'protects' his eroticism with the biblical account of Isaac in intimate congress with his much younger wife Rebecca, his protégé ensconces his own homoerotic scene of an older Apollo and his young amour in mythological narrative. Michelangelo spent eight years executing his Florentine Pietà (1547-53)[15] for his own tomb. The growing popularity of this sexual idiom, as Steinberg and I concur, may well have been the catalyst for him taking a hammer to Christ's left leg.[16] This hypothesis gains considerable currency when it is found to be used in a homoerotic context. Michelangelo's repression of his own sexual ambivalence has already been the subject of exhaustive scholarly psychoanalysis. But the plausibility that a nuptial portent originally conceived to symbolise Christ's mystic espousal of Mary provoking destructive rage in its tormented creator because it was perhaps too invoking of his own battle to suppress sinful pederastic temptation invites further discourse.

Further indications that Romano intended Apollo and Cyparissus to be understood as a metaphoric analogy for rites of initiation and passage into manhood are revealed in the phallic symbolism of the two bows. The one Hyacinth holds is the weapon which killed, albeit accidentally, his stag. It stands proud and upright because, in accordance with recorded rites of passage associated with many cultures, a juvenile becomes a man when he is deemed mature enough to join his peers on hunting exploits. Now he has successfully passed this initiation rite Cyparissus is deemed worthy and permitted to pass onto the rank of adult male citizen. Hence, rather than seeing their kiss as erotic passion alone, perhaps it is also one of farewell. Romano juxtaposes the proud display of this manly bow with the redundant discarded one in the central foreground. The cello positioned precariously leaning as if abandoned against the rock, rather than that of the female figure, replicates the graceful curves of idealised womanly form. Together with the discrete serpent head engraved on its neck referencing female temptation, this forsaken instrument with its strongly suggestive feminine characteristics lies neglected by the presently preoccupied juvenile Cyparissus. As if to emphasise that passivity was only acceptable at a certain stage in a man's life, the phallic symbolism of the temporarily forsaken bow evokes a certain reassurance that once transformation into adulthood as an active procreative being is complete Cyparissus will return his attention to these more 'natural' pursuits in order to fulfil his own patriarchal and pedagogical duty. The pivotal role played by the cypress in the Ovidian narrative is recalled by the tree Romano places as the central axis of the composition; thus balancing norm against transgression. To the right of the scene the intimate embrace of the two male paramours signify his adolescent passive exploits, but soon he will step into the manly sphere of active adulthood that awaits him on the other side. With his astute compositional strategy Romano cogently

ISBN 978-0-9566139-1-2

underscores how sexual experiences between males in the Renaissance were only tolerated if they were successive and cyclical and not seen as an alternative to marriage and producing a family.

CONCLUSION

Both these works conform to the contemporary culturally defined sexual and social roles relating to sexuality, gender and identity that operated in Renaissance Italy despite their homoerotic subject matter. Not only do they challenge notions that heterosexuality was normative and universal in this period, but as mythological representations of the superordinate Apollo and his subordinate beloveds they capture the gradations in social status and power differentials that dominated their age. The understanding and contextualizing of the dynamics of same-sex eroticism and sociality, as represented in the artistic production of the late medieval and early modern period, is an important but so far relatively neglected area of study within the field of art historical discourse. This paper has aimed to demonstrate the ways in which the prevalence and variety of social and sexual behaviours and experiences between men in the Renaissance, as well as the psychological meanings, patterns and identities assigned to those acts, found expression in the visual art of the period. It is hoped that by offering new interpretations of these remarkable mythological exemplars of Renaissance homoerotic imagery a small contribution has been made in edging this fascinating subject closer to the intellectual mainstream.

REFERENCES

[1] Ovid, *Metamorphoses*, trans. A.D. Melville, Oxford World Classics, (1987) Oxford: Oxford University Press. Pp 249-272.
[2] Rocke, M., Forbidden Friendships: Homosexuality and Male Culture in Renaissance Florence. (1999), Oxford: Oxford University Press, pp. 78-109.
[3] Benvenuto Cellini, *Apollo and Hyacinth*, 1545-48, Marble, Museo Nazionale del Bargello, Florence.
[4] Giulio Romano, *Apollo and Cyparissus*, c.1523, Pen and ink with wash, National Museum, Stockholm.
[5] Cole, M. *Cellini and the Principles of Sculpture.* (2005) Cambridge: Cambridge University Press, pp. 83-85.
[6] Talvacchia, B. *Taking Positions: On the Erotic in Renaissance Culture.* (1999) Princeton: Princeton University Press, p.134.
[7] Apollo Belvedere, c.130-140 A.D., Rome, Vatican Museums.
[8] Saslow, J., *Ganymede in the Renaissance: Homosexuality in Art and Society*, Yale University Press, New Haven and London, 1986, p. 152.
[9] Cellini, B. *Life.* trans. Bull, (1956) London: Penguin Books, p.338.
[10] Steinberg, L. *The Sexuality of Christ in Renaissance Art and in Modern Oblivion.* (1983) Chicago: The University of Chicago Press
[11] Giovanni Cariani, *Madonna and Child with Donor*, 1520, Bergamo, Accademia Carrara; Domenico Ghirlandaio, *Adoration of the Magi*, 1497, Galleria degli Uffizi, Florence.
[12] Giulio Romano, *I Modi: Position 7*, 1524, engraving by Marcantonio Raimondi, 1527. Private Collection, Geneva.
[13] Giulio Romano, *Madonna and Child*, 1522-23, Oil on Panel, Galleria Nazionale, Palazzo Barberini, Rome.
[14] Rafaello Sanzio, *Isaac and Rebecca spied upon by Abimelech*, 1518-19, Fresco, Palazzi Pontifici, Vatican.
[15] Michelangelo Buonarroti, *Pietà*, 1547-53, Museo dell'Opera del Duomo, Florence.
[16] Steinberg, L., *Michelangelo's Florentine Pietà: The Missing Leg*, The Art Bulletin, Vol. 50, No. 4 (Dec., 1968), pp. 343-353, http://www.jstor.org/stable/3048570

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

THE DUAL TRAJECTORIES OF CHRONIC PAIN AND FALLING IN COMMUNITY DWELLING OLDER ADULTS – TRANSGRESSING CONCEPTUAL AND PARADIGMIC BOUNDARIES

Pamela IRWIN

PhD Student, University of Oxford, Oxfordshire, UK
pamela.irwin@sociology.ox.ac.uk

Keywords: Ageing, Long-term health conditions, Chronic pain, Falls

ABSTRACT

The parallel trajectories of community dwelling older adults experiencing chronic pain and/or falls usually represent a downward cascade of events to ultimately create or increase social isolation and societal exclusion. As both chronic pain and falling are characterised by a complex array of contributing and potentially confounding factors, a multidimensional, multidisciplinary investigative approach is required to tease out complicated relationships. For instance, although a preliminary literature review by Verala-Burstein and Miller uncovered a possible link between the incidence of chronic pain and the risk of falling, this association has yet to be fully explored. Similarly, while recent longitudinal analyses have enabled sophisticated models of the ageing process, a disjuncture remains between these more 'empirical' profiles and complementary qualitative narratives describing the real-world impact of chronic pain and falls in older people's lives. Recent studies, such as Mahler and Sarvimäki's interpretative phenomenological research into the meaning of falling, and Frank's autobiographical stories about living with chronic pain, provide rich contextual pictures of the day-to-day reality of older persons' lives. My paper '*The dual trajectories of chronic pain and falling in community dwelling older adults*' reflects the interdisciplinarity of the '*Diverse Engagement: Drawing in the Margins*' conference, and as such, will adopt an holistic biopsychosocial approach to population demographics, social gerontology, sociology, health care, and predictive statistical modelling.

DIVERSE ENGAGEMENT – drawing in and across the margins

Recent probabilistic forecasting in population demography clearly signals a global increasing and accelerating ageing phenomenon.[1] Modern conceptions of ageing meld age-related biological and physical changes with culturally and socially constructed normative representations of age identity,[2] and as such, stretch the classic demographic determinants – fertility, migration and mortality.[3]

When new indicators capturing changes in longevity are incorporated into conventional models of ageing,[4] the emergent profile underscores both the criticality of healthy life expectancy, as well as its corollary, the challenge of countering morbidity. These population-level social determinants of health[5] then funnel into corresponding finer grained instruments with functional utility at societal and individual levels. Composite measures such as quality of life indices[6] also enable researchers to unpick the trans-dimensional patterns embedded in many long-term health problems.

From a lifespan perspective, the supplementary lens of older age represents an additional layer of complexity with respect to health,[7, 8] giving rise to conceptual, methodological and operational issues. Challenges extend

ISBN 978-0-9566139-1-2

from difficulty integrating social, emotional and economic factors into population studies of ageing and health[9] to reaching definitional consensus for some geriatric syndromes such as self-neglect[10 11] and sarcopenia.[12]

Many of these sticking points symbolise the unresolved areas of contention that characterise this arena – for example, well-being in later life versus age-related decline – each with its associated specialisms. Consequently, gerontology is allied with the broadly conceived social sciences construct, 'successful ageing',[13, 14] while management of 'geriatric syndromes' is linked with the medically grounded discipline of geriatrics. Further specialisation has since created new sub-disciplines; with sociology evolving into social gerontology and social neuroscience, and medicine, bio-epidemiology, among others.

This paper will 'diversely engage' with these contested areas via a detailed review of two age related chronic health conditions, and conclude by proposing that 'drawing in and across the margins' is a potentially unifying and integrative mechanism across the life span.

Defining the margins

While the parallel trajectories of chronic pain and/or falls in community dwelling older adults are usually represented as a downward cascade of events creating and/or increasing social isolation and exclusion,[15] when examined more closely, some interesting deviations are evident. A preliminary survey of the extensive research literature pertaining to falls suggests a strong, 'objective' clinical focus, whereas elderly chronic pain is significantly under-researched[16] and the contributing variables are often viewed more 'subjectively'.[17] These apparent paradigmic differences inform complementary approaches to a more detailed analysis of the literature. A systematic review (meta-analysis) aligns with a 'hard', evidence-based[18] orientation and is the tool of choice to analyse falls research. Alternatively, a systematic narrative synthesis[19] provides a closer fit with the 'softer' style of pain-focused literature.

Pain of more than six months duration is defined as chronic (persistent or recurrent), and is estimated to affect more than 50% of older people who live in the community.[20] In the elderly, pain tends to be constant, moderately to severely intense, and last for several years.[21] Conversely, the pattern of falling in community dwelling adults over 65 years old, is episodic and cumulative, with approximately 28-35% of this cohort sustaining one fall per year.[22]

Although the prevalence rate of falling is lower than the incidence of chronic pain, falls may lead to fractures, with concomitant threats to morbidity and mortality.[23] Direct and indirect falls-related health care costs are also significant. A comparison of United States of America medicare fallers and non-fallers, found that sustaining one fall increased aggregated costs by 29%; this figure escalated to 79% for recurrent falls.[24] These statistics have precipitated numerous investigations into the complicated web of factors contributing to falls in the elderly, ranging from the seminal study by Tinetti[25], to sophisticated predictive models by Campbell and Robertson.[26] The collective outcomes of this work suggest that the risk of falling coalesces into four dimensions - biological, behavioural, environmental and socioeconomic factors.[27] When differentiated into mutable and non-mutable components, they form the backbone of many falls prevention programmes.[28]

Chronic pain in the elderly is also multifactoral and multifocal, but has not received the same research attention. Despite Kahana, Kahana, Namazi, Kercher and Stange's[29] preliminary examination of the cascade of ageing, pain, long-term illness, social disability, and psychological distress, comparatively little more is known about the personal, social and economic consequences of recurrent or chronic pain in older adults.[30] This deficit is attributed to various causes, including the under-recognition of persistent pain, with symptoms often attributed to normal processes, other co-morbid and chronic conditions,[31] and ageism by both health care practitioners and the elderly themselves.[32] Inconsistencies inherent in the measurement of pain in older adults amplify this position, where '*systematic efforts have only just started to focus on the reliability and validity of common pain assessment tools when used in older populations*'.[33] Guidelines issued by the British Pain Society and British Geriatrics Society in 2007 also reinforce the need for a multifaceted, comprehensive assessment of pain in the aged.[34] Many studies investigating the management of chronic pain in older people are severely impeded by stringent design parameters - in effect, excluding the population of interest from participating in research.[35] To compensate for this lack of rigour, Karp, Shega, Morone and Weiner[36] advocate pain care plans that individualise age-related and

ISBN 978-0-9566139-1-2

pathological changes, pharmacokinetic and pharmacodynamic effects, caregiving burden and distress, and additional socio-economic issues.

By contrast, research into falls in the elderly is centred on increased evidence and robustness in the form of randomized controlled trials, systematic reviews, and clinical guidelines. As the major determinants of falling are now relatively well documented,[37] major research efforts are aimed at uncovering the most effective falls prevention[38] and/or intervention strategies.[39]

However, despite the accumulating evidence supporting the success and benefit of multidimensional chronic pain[40] and falls[41, 42] interventions, the outcome path for many elderly people experiencing chronic pain and/or falls, remains a negative downward spiral. This worrying scenario invites an exploration of supplementary explanatory influences.

Drawing in the margins

A recent study exploring well-being in later life[43] is designed to plug information gaps in older people's experiences of ageing, quality of life, expectations and emotional health. It builds on previous work suggesting that quality of life in old age is differentially valued and filtered through meaning, and the participant's 'lived' experience is richer and more textured than the standardised norms.[44] When contextualised as chronic health conditions, Frank[45] and Van Heut, Innes and Whiteford's[46] biographical stories about living with chronic pain provide thick pictures of the day-to-day reality of older persons' lives, while Mahler and Sarvimäki's[47] interpretative phenomenological study into the meaning of falling is a relatively uncommon adjunct to typical falls research.

The significance of '*meaning-making*'[48] in older people's lives highlights the disjuncture between empirical profiles and complementary qualitative narratives describing the real-world impact of chronic health conditions. While the social epidemiology of ageing indicates that community dwelling elders who live alone are more likely to be older, female, and suffering from chronic physical or mental health problems, more research is needed to unscramble the social isolation-health status dyad.[49] This omission is particularly disturbing in the context of Barnes, Blom, Cox, Lessof and Walker's[50] analysis of social exclusion in older English people, as well as contemporaneous sociological research into the importance of close relationships and social networks as a counter to isolation in the elderly. Yet, little '*interest has been shown in understanding the meaning of such relationships and how they are mediated and shaped by both the broad socio-spatial context and the personal biographies and experiences of the protagonists.*'[51] When the impact of this social isolation is harmonised with older adults experiencing chronic pain and/or falls, the nature of the synthesis is also an under-researched enigma.

Similarly, even though reports of chronic pain and falling are both very frequent in older adults, the influence of pain on the risk of falling, or vice versa, has not been established. Despite initial progress in this area[52], it is still unknown whether the incidence of chronic pain generally will increase or decrease the risk of falling in elderly people. Arguments can be proposed for both perspectives: a person in pain may 'guard' their physical actions and this could protect against falling, or alternatively, pain may negatively alter normal movement patterns and add to the risk of falls.

In addition, although falling can produce painful injuries, most of the literature describing the consequences of falls relates to acute rather than chronic or recurrent pain. Gold and Roberto[53] attribute this oversight to the multifactoral and multidisciplinary nature of chronic pain. As such, chronic pain does not conform to accepted professional demarcations, and related research findings are dispersed among many specialty journals. An analogous pattern occurs with the falls literature.

Engagement across the margins

At this stage, it is not clear if this array of contributing and/or confounding factors will facilitate or hinder cross-disciplinary collaborations clustered around chronic pain and falling in older adults. With respect to transgressing and extending the traditional boundaries of population demographics, predictive statistical modeling, social gerontology, life course sociology and chronic health conditions though, some encouraging trends are emerging.

ISBN 978-0-9566139-1-2

Despite the biomedical model's continued dominance of geriatrics,[54] an integrated biopsychosocial structure is now widely employed across the spectrum of health and social care theory and practice. Originally proposed to articulate the dynamic interplay between biological, psychological and social factors and health outcomes, the biopsychosocial model was not developed to explain causal influences or interactions.[55] However this methodological limitation can be minimized with large scale longitudinal research designs and statistical analyses to tease out complex relations through predictive inferences about the cumulative temporal, age-period-cohort, and social processes that change over a life course.[56] A multiple methods approach also enhances interpretative capacity, since mixed methods generally involve a synthesis of quantitative and qualitative perspectives and/or intellectual and practical orientations to inform different 'sense-making'.[57] These early signposts point to an 'ideal' holistic framework that will map temporal and cumulative life course trajectories ('upstream determinants) against potential 'downstream' outcomes; accommodate critical incidents (transitions) while accounting for age, period and cohort effects; and incorporate demographic and epidemiological determinants with lived experience.

Even though the challenge of translating the multidimensional complexity of both chronic pain and the risk of falling into an accessible, parsimonious model is extremely daunting, 'diversely engaging in and across the margins' does offer a potential unifying and heuristic mechanism for demystifying these age related chronic health conditions across the lifespan.

REFERENCES

[1] Leeson, G. *The ageing of Europe's populations – demographic divergence or convergence?* (2005). Dunblane, Scotland: Scottish Executive/European Economic and Social Committee.
[2] Diehl, M.K., and Wahl, H-W. "Awareness of age-related change: examination of a (mostly) unexplored concept." *Journal of Gerontology: Social Science*s 10 (2009): 1093/geronb/gbp110. (06 April, 2010).
[3] Harper, S. *Ageing societies.* (2006). London: Hodder Arnold.
[4] Lutz, W., Sanderson, W., and Scherbov, S. "The coming acceleration of global population ageing." *Nature* 451 (07 February, 2008): 716-719.
[5] Commission on Social Determinants of Health. *Closing the gap in a generation: health equity through action on the social determinants of health. Final Report of the Commission on Social Determinants of Health.* (2008). Geneva: World Health Organisation.
[6] Hildon, Z., Montgomery, S.M., Blane, D., Wiggins, R.D., and Netuveli, G. "Examining resilience of quality of life in the face of health-related and psychological adversity at older ages: what is "right" about the way we age?" *The Gerontologist* 50: 1 (2009): 36-47.
[7] Denton, F.T., and Spencer, B.G. "Chronic health conditions: changing prevalence in an aging population and some implications for the delivery of health care services." *Canadian Journal on Aging* 29: 1 (2010): 11-21.
[8] Freeman, V.A., Schoeni, R.F., Martin, L.G., and Cornman, J.C. "Chronic conditions and the decline in late-life disability." *Demography* 44: 3 (August 2007): 459-477.
[9] Steptoe, A. "Psychosocial biomarker research: integrating social, emotional and economic factors into population studies of aging and health." *Social Cognitive and Affective Neuroscience Advance Access*: doi:10.1093/scan/nsq032. (12 April, 2010): 1-8.
[10] Iris, M., Ridings, J.W., and Conrad, K.J. "The development of a conceptual model for understanding self-neglect." *The Gerontologist* 50: 3 (2010): 303-315.
[11] Paveza, G., VandeWeerd, C., and Laumann, E. "Elder self-neglect. A discussion of a social typology." *Journal of American Geriatrics Society* 56: 52 (2008): S272-S275.
[12] Cruz-Jentoft, A.J., Baeyens, J.P., Bauer, J.M., Biore, Y., Cederholm, T., Landi, F., Martin, F.C., Michel, J-P., Rolland, Y., Schneider, S.M., Topinkova, E., Vandewoude, M., and Zamboni, M. "Sarcopenia: European consensus on definition and diagnosis." *Age and Ageing Advance Access*: doi: 10.1093/ageing/afq034 (06 May, 2010): 1-12.
[13] McLaughlin, S.J., Connell, C.M., Heeringa, S.G., Li, L.W., and Roberts, J.S. "Successful aging in the United States: prevalence estimates from a national sample of older adults." *Journal of Gerontology: Social Sciences* 65B: 2 (2010): 216-266
[14] Pruchno, R.A., Wilson-Genderson, M., Rose, M., and Cartwright, F. "Successful aging: early influences and contemporary characteristics." *The Gerontologist Advance Access*: doi: 10.1093/geront/gnq041 (01 June 01, 2010).
[15] Barnes, M., Blom, A., Cox, K., Lessof, C. and Walker, A. *The social exclusion of older people: evidence form the first wave of the English Longitudinal Study of Ageing (ELSA), Final Report.* (2006). London: Social Exclusion Unit, Office of the Deputy Prime Minister. HMSO.
[16] Gibson, S.J. "Older people's pain." *Pain: clinical updates* 14 (2006): 1-4.
[17] Sim, J., and Smith, M.V. "The sociology of pain." in Sally French and Julius Sim, Eds, *Physiotherapy: a psychosocial approach.* 3rd Ed. (Edinburgh: Elsevier Butterworth Heinemann, 2005), 117-139.

ISBN 978-0-9566139-1-2

[18] Muir Gray, J.A. *Evidence-based healthcare. How to make health policy and management decisions*. 2nd Ed. (2006). Edinburgh: Churchill Livingstone.
[19] Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., and Duffy, S. *Guidance on the conduct of narrative synthesis in systematic reviews*. (2006). ESRC Methods Programme.
[20] Helme, R.D., and Gibson, S.J. "The epidemiology of pain in older people." *Clinics of Geriatric Medicine* 17 (2001): 417-431.
[21] Brattberg, G., Parker, M.G., and Thorslund, M. "A longitudinal study of pain: reported pain from middle age to old age." *Clinical Journal of Pain* 13 (1997): 144-149.
[22] World Health Organisation. *WHO global report on falls prevention in older age*. (2007). Geneva: World Health Organisation.
[23] World Health Organisation. *WHO global report on falls prevention in older age*. (2007). Geneva: World Health Organisation.
[24] Shumway-Cook, A., Ciol, M.A., Hoffman, J., Dudgeon, B.J., Yorkston, K. and Chan, L. "Falls in the medicare population: incidence, associated factors, and impact on health care." *Physical Therapy* 89: 4 (2009): 1-9.
[25] Tinetti, M.E. "Preventing falls in elderly persons." *The New England Journal of Medicine* 348: 1 (2003): 42-49.
[26] Campbell, A.J., and Robertson, M.C. "Rethinking individual and community fall prevention strategies: a meta-regression comparing single and multifactoral interventions." *Age and Ageing* 36 (2007): 656-662.
[27] World Health Organisation. *WHO global report on falls prevention in older age*. (2007). Geneva: World Health Organisation.
[28] Lamb, S.E., Fisher, J.D., Gates, S., Potter, R., Cooke, M.W., and Carter, Y.H. "A national survey of services for the prevention and management of falls in the UK." *BMC Health Services Research* 8 (2008): 233-41.
[29] Kahana, B., Kahana, E. Namazi, K., Kercher, K., and Stange, K. "The role of pain in the cascade from chronic illness to social disability and psychological distress in late life." in D.I. Mostofsky and J. Lomranz. *Handbook of Pain and Aging,* (NY: Plenum Press, 1997), 185-206.
[30] Gibson, S.J. "Older people's pain." *Pain: clinical updates* 14 (2006): 1-4.
[31] Karp, J.F., Shega, J.W., Morone, N.E., and Weiner, D.K. "Advances in understanding the mechanisms and management of persistent pain in older adults." *British Journal of Anesthesia* 101: 1 (2008): 111-120.
[32] Schofield, P. "It's your age: the assessment and management of pain in older adults." *Continuing Education in Anesthesia, Critical Care and Pain* 10: 3 (2010): 93-95.
[33] International Association for the Study of Pain. "Older people's pain." *Pain: Clinical Updates* XIV: 3 (2006), p.2.
[34] British Pain Society and British Geriatrics Society. *Guidance on the assessment and treatment of pain in older people.* (2007). London: British Pain Society.
[35] Lussier, D. "Pain in older persons: urgent need for research." in *Newsletter of the IASP Special Interest Group on Pain in Older Persons.* (International Association for the Study of Pain, June 2007).
[36] Karp, J.F., Shega, J.W., Morone, N.E., and Weiner, D.K. "Advances in understanding the mechanisms and management of persistent pain in older adults." *British Journal of Anesthesia* 101: 1 (2008): 111-120.
[37] Rubenstein, L.Z. "Falls in older people: Epidemiology, risk factors and strategies for prevention." *Age and Ageing* 35-S2 (2006): ii37-ii41.
[38] Markle-Reid, M., Brown, G., Gafni, A., Roberts, J., Weir, R., Thabane, L., Miles, M., Vaitonis, V., Hecimovich, C., Baxter, P., and Henderson, S. "The effects and costs of a multifactoral and interdisciplinary team approach to falls prevention for older home care clients 'at risk' for falling: a randomized controlled trial." *Canadian Journal on Aging* 29: 1 (2010): 139-161.
[39] Gillespie, L.D., Robertson, M.C., Gillespie, W.J., Lamb, S.E., Gates, S., Cumming, R.G., and Rowe, R.H. "Interventions for preventing falls in older people living in the community (review)." *The Cochrane Collaboration.* Issue 4. (2009).
[40] Gagliese, L. and Melzack, R. "Pain in the elderly." in Stephen B. McMahon and Martin Koltzenburg, eds. *Wall and Melzack's Textbook of Pain*. 5th Edition, (Philadelphia: Elsevier Churchill Livingstone, 2006), 1169-1179.
[41] Center for Evidence-Based Practice. *A systematic review on prevention of falls in community-dwelling elderly persons.* (2009) Pittsburg PA: University of Pennsylvania Health System
[42] Gillespie, L.D., Robertson, M.C., Gillespie, W.J., Lamb, S.E., Gates, S., Cumming, R.G., and Rowe, R.H. "Interventions for preventing falls in older people living in the community (review)." *The Cochrane Collaboration.* Issue 4. (2009).
[43] McCormick, J., Clifton, J., Sachrajda, A., Cherti, M. and McDowell, E. *Getting on: well-being in later life.* (2009). London: Institute for Public Policy Research.
[44] Borglin, G., Edberg, A-K., and Hallberg, I.R. "The experience of quality of life among older people." *Journal of Aging Studies* 19 (2005): 201-220.
[45] Frank, A.W. "Asking the right question about pain: narrative and phronesis." *Literature and Medicine* 23: 2. (Fall 2004): 209-225.
[46] Van Heut, H., Innes, E., and Whiteford. G. "Living and doing with chronic pain: narratives of pain program participants." *Disability Rehabilitation* 31: 24 (2009): 2031-2040.
[47] Mahler, M., and Sarvimaki, A. "Indispensible chairs and comforting cushions – falls and the meaning of falls in six older persons lives." *Journal of Aging Studies* 24 (2010): 88-95.
[48] Coakley, S. "Afterword" in Sarah Coakley and Kay Kaufman Shelemay, eds. *Pain and its Transformations: the Interface of Biology and Culture.* (Boston: Harvard University Press, 2008): 423-427.
[49] Victor, C., Scambler, S., and Bond, J. *The social world of older people.* (2009). Maidenhead, Berkshire: McGraw Hill Open University Press.

ISBN 978-0-9566139-1-2

[50] Barnes, M., Blom, A., Cox, K., Lessof, C. and Walker, A. *The social exclusion of older people: evidence form the first wave of the English Longitudinal Study of Ageing (ELSA), Final Report.* (2006). London: Social Exclusion Unit, Office of the Deputy Prime Minister. HMSO.

[51] Victor, C., Scambler, S., and Bond, J. *The social world of older people.* (2009). Maidenhead, Berkshire: McGraw Hill Open University Press. p.8-9.

[52] Varela-Burstein, E., and Miller, P.A. "Is chronic pain a risk factor for falls among community dwelling elders?" *Topics in Geriatric Rehabilitation* 19: 2 (2003): 145-159.

[53] Gold, D.T., and Roberto, K.A. "Correlates and consequences of chronic pain in older adults." *Geriatric Nursing* 21: 5 (2000): 270-273.

[54] Fava, G.A., and Sonion, N. "The biopsychosocial model thirty years later." *Psychotherapy and Psychosomatics* 77 (2008): 1-2

[55] Borrell-Carrio, F., Suchman, A.L. and Epstein, R.M. "The biopsychosocial model 25 years later: principles, practice and scientific inquiry." *Annals of Family Medicine* 2: 6 (2004): 576-582.

[56] Hofer, S.M., and Piccinin, A.M. "Toward an integrative science of life-span development and aging." *Journal of Gerontology: Psychological Sciences*: 10.1093/geronb/gbq017 (16 April 06, 2010).

[57] Burke Johnson, R., Onwuegbuzie, A.J., and Turner, L.A. "Towards a definition of mixed methods research." *Journal of Mixed Methods Research* 1: 2 (2007): 112-133.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

CONSEQUENCE OF PRECIPITATION ON PATHOGENS, MICROBIAL WATER QUALITY AND DISEASE OUTBREAKS

Tanvir ISLAM

Department of Civil Engineering, Water and Health Research Centre,
University of Bristol, Bristol BS8 1UB, United Kingdom
tanvir.islam@bristol.ac.uk

Keywords: Pathogens, Microbiological water quality, Rainfall, Precipitation, Water management

ABSTRACT

The presence of pathogenic microorganisms in surface and groundwater system is an imperative issue. This paper reviews the significance of precipitation on pathogens affecting the water quality in microbiological perspective. There is a strong association of precipitation with microorganisms and their fate and transport in the environment. Level of water quality severely declines due to rainfall. Contamination becomes even worse during extreme events with the increase of microbial loads. Numerous disease outbreaks are rottenly linked with the consequence. A good framework is necessitated considering the meterological condition in water management and health surveillance system preparing the professionals in advance.

INTRODUCTION

The occurrence of water contamination by pathogenic microorganisms has long been a concern. Water is subject to the risk of microbiological contamination in many circumstances and sources. Both point and non point sources of pollution could link to pathogenic contamination of water. Humans and animals to inadequately treated sewage discharge[1] to wastewater treatment plants,[2] all are an obvious high risk source of pathogens. Agricultural practice is cited also as a major source of pathogenic microorganisms in surface and groundwater system.[3, 4] On site sanitation is another risk to contamination.[1, 5, 6] During period of precipitation, pathogens may be dispersed from runoff and may cross the threshold into drinking water sources. However, no single indicator or simple hydrological index is entirely suitable to comment about the association.[7] In a distribution system, occurrence of microorganisms depends upon a complex interface of physical, chemical, operational and engineering parameters.[8]

The entry of pathogenic microorganisms into drinking water sources poses a great risk to human health.[9] Good understanding of fate and transport of pathogenic pollutants associated with precipitation is important to tackle the risk. Probing relationships between precipitation and pathogenic microorganisms could help improve our understanding of the processes involved.[7] Good firm understanding is required on microorganism distribution in the environment for a sustainable water management.

CONSEQUENCE IN SURFACE WATER

The mechanism of pathogens movement in surface water is dependent on assorted variables.[10] Pathogens vary considerably with respect to a variety of factors that manipulates their fate and transport in the environment.[11-15] During precipitation event, hydrology thus water flow changes what is believed to control the transport and fate of microorganisms in water courses.[16-19] Increment of hydraulic flow could amplify the concentration of microorganisms.[20-22]

ISBN 978-0-9566139-1-2

Numerous researchers have noted the consequence of rainfall on microorganisms and water quality. Ferguson et al.[23] examined water quality in an estuarine system and found that rainfall affected the water quality with a significant increase in the concentration of fecal coliforms, fecal streptococci, *Clostridum perfringens* spores, F-RNA bacteriophage, *Aeromonas* spp., *Giardia* and *Cryptospordium* spp. However, result was also influenced by sewage overflow as the study encompassed the tidal regions of a river which receives inputs of urban runoff from surrounding metropolitan areas and during wet weather, may also collect sewage contributions. To determine the variables, statistical models were generated including precipitation in the previous 72 h and its association was observed in the water column with the significant increases in the concentration of microorganisms. In an intensive study in New Zealand between 1997 to 2006, Britton et al.[24] analyzed a spatial relationships between cryptosporidiosis, giardiasis notifications and climatological average precipitation by negative binomial regression. They discovered a positive relationship between the parameters. Atherholt et al.[25] evaluated the relationship between *Giardia* and *Cryptosporidium* concentrations and other bacteriological, water quality, meterological parameters. The investigation was conducted in the Delaware River Basin (DRB) in 1996. The year was a record breaking year of DRB and it experienced its wettest year in 102 years of record. Sample sets were collected for each season and Spearman rank method was used to inspect the parasite concentrations. Seasonal difference in parasite concentrations related to meterological factor was observed with a positive correlation ($p<0.05$) of *Giardia* or *Cryptosporidium* and 15 other parameters. Golder and Banarjee[26] carried out an econometric analysis of determinants of water quality in Indian rivers and estimated an ordered probit model for the explanation of water quality variations. They did not find any relation of water quality with industrialization, irrigation intensity and fertilizer use, however, an optimistic relation with rainfall was identified. Egwari and Aboaba [27] concluded rainfall accentuated the impact on the bacteriological quality of domestic water supplies in Lagos, Nigeria. Stukel et al.[28] conducted a longitudinal study of rainfall and coliform contamination in small community drinking water system. Samples were tested for a year and data with precipitation measurements were analysed by statistical methodology for repeated binary responses. After analyzing, they reported that the odds of total coliforms were positively increased. [29] exhibited that concentrations of *Cryptosporidium* were lower during dry period as compared to high runoff period. Mons et al.[30] found significant association of enterococci counts and rainfalls with *Giardia* concentration.

CONSEQUENCE IN GROUND WATER

Distribution of pathogens in groundwater due to rainfall is a complex phenomenon. Several factors control the transport of pathogens during the precipitation event. It is well fact that moisture increases significantly during precipitation period, which plays an important role to groundwater contamination. Many researchers have implied that the principal factor affecting the survival of enteric bacteria in the soil system is the moisture status.[18, 31-38] With the presence of moisture, pathogens transport from the surface into subsurface. Within the thin film of moisture, the organisms contact with the particle surface and thus, soil moisture influences the persistence of microorganisms.[36] During periods of high recharge, the intergrannular spaces in the unsaturated zone become waterlogged thus provides a hydraulic pathway for the pathogen transportation. Indeed, following rainfall, springs recharge rapidly, ultimately leads to bacteriological contamination.[36]

Microbiological water quality in groundwater considerably falls during rainfall period. Godfrey et al.[39] presented evidence of direct pulse response of groundwater pollution events to rainfall. They analyzed different information including wells water quality over a 12 month period, historical precipitation from the previous 8 years, water table depth etc. and confirmed the increase of presumptive thermotolerant coliforms and enterococci bacteria with a direct pulse response. Hagedorn et al.[40] found highest *Escherichia coli* after a rise in the water table following major rainfall events. Tate[38] found *Escherichia coli* survival to be greatest in organic soils under flooded condition. Barrell and Rowland[41] observed an increase of 10 orders of magnitude of fecal contamination following the onset of rains for a 6 days period in Gambia. Further studies by Barrett et al.[42] and Howard et al.[6] performed in Kampala, Uganda showed a significant deterioration in microbial groundwater quality within 12 hours of a rainfall event.

ISBN 978-0-9566139-1-2

PRECIPITATION INTENSITY

Several investigations have been carried out to look at pathogenic concentration after rainfall and it is found that water quality severely declines following heavy rainfall[30, 43]. Kay et al.[44] monitored the bacteriological water quality of seven large commercial private water supplies in the United Kingdom and poor water quality was typically observed following heavy rainfalls. However, very few studies have been conducted to date which clearly distinguishes the precipitation intensity to microbiological water quality. Kistemann et al.[45] conducted such a study by investigating colony, *Escherichia coli*, coliform, fecal streptococcal, and *Clostridium perifingens* counts. They quantified and compared the microbial loads of drinking water reservoir tributaries during extreme rainfall events with regular events. They monitored the microbial loads of three tributaries for a period of 14 month under different runoff conditions, which were then analyzed by statistical methods. Outcome of the study clearly indicated that substantial shares of the total microbial loads produced from extreme rainfall and runoff event.

DISEASE OUTBREAKS

Several waterborne disease outbreaks have been reported during the past few decades, which had a direct or indirect relationship with precipitation. Nichols et al.[46] showed the evidence that rainfall precede many waterborne outbreaks. A study conducted by Curriero et al.[17] in the United States from 1948 to 1994 confirmed the implication, and fifty-one percent of waterborne disease outbreaks were preceded by precipitation events above the 90th percentile (P =.002), along with 68% by events above the 80th percentile (P=.001). Thomas et al.[47] described the role of high impact weather events with incidence and distribution of disease outbreaks in Canada. Using the case-crossover methodology, they examined extreme rainfall and spring snowmelt in association with 92 Canadian waterborne disease outbreaks between 1975 and 2001. Accumulated rainfall percentiles were associated with outbreak risk and were considered as the contributing factors. The Walkerton is another example of disease outbreaks' association with rainfall.[48] Study confirmed *Escherichia coli* O157:H7 and *Campylobacter* outbreak occurred several days after a heavy rainfall. Jean et al.[43] tested enterovirus after a heavy rainfall in Southern Taiwan using the reverse transcription-polymerase chain reaction (RT-PCR) assay. Following detection, the logistic regression model revealed a statistical links between the rainfall rate and observed enterovirus infection case. Hicks et al.[49] demonstrated using negative binomial model that Legionellosis incidence increased during periods of increased rainfall. Hashizume et al.[50] concluded that rainfall in the preceding weeks increased the number of non-cholera diarrhoea cases both above and below a threshold level in Bangladesh.

CONCLUSION

It is clearly understood that precipitation significantly associates with water pollution and waterborne epidemics. However, meteorological conditions are often not coupled with water quality monitoring and there is no apparent guideline for the matter. A good framework is eventually needed by integrating the meterological condition with water quality monitoring for better water management. Indeed, weatherise condition could give a good indication of the possible challenge, thus could avoid the likely outbreaks. Considerable loss can be prevented by better policy making and management. A weather forecast system can be implemented assimilating water and health monitoring program which could prepare the water and health professionals in advance.

REFERENCES

[1]M. Barrett, G. Howard, S. Pedley, R. Taylor and M. Nalubega, "A comparison of the extent and impacts of sewage contamination on urban groundwater in developed and developing countries", *Schriftenr Ver Wasser Boden Lufthyg* 105 (2000): 179-85.

[2]M. Dechesne and E. Soyeux, "Assessment of source water pathogen contamination", *Journal of Water and Health* 5 (2007): 39-50.

[3]R. Khaleel, K. R. Reddy and M. R. Overcash, "Transport of potential pollutants in runoff water from land areas receiving animal wastes - a review", *Water Research* 14 (1980): 421-436.
[4]A. R. Tiedemann, D. A. Higgins, T. M. Quigley, H. R. Sanderson and D. B. Marx, "Responses of fecal-coliform in streamwater to 4 grazing strategies", *Journal of Range Management* 40 (1987): 322-329.
[5]M. H. Barrett and A. G. Howard, "Urban groundwater and sanitation - Developed and developing countries", in K. W. F. Howard and R. G. Israfilov, eds., *Current Problems of Hydrogeology in Urban Areas, Urban Agglomerates and Industrial Centres* (Dordrecht, 2002), 39-56.
[6]G. Howard, S. Pedley, M. Barrett, M. Nalubega and K. Johal, "Risk factors contributing to microbiological contamination of shallow groundwater in Kampala, Uganda", *Water Research* 37 (2003): 3421-3429.
[7]G. Wilkes, T. Edge, V. Gannon, C. Jokinen, E. Lyautey, D. Medeiros, N. Neumann, N. Ruecker, E. Topp and D. R. Lapen, "Seasonal relationships among indicator bacteria, pathogenic bacteria, Cryptosporidium oocysts, Giardia cysts, and hydrological indices for surface waters within an agricultural landscape", *Water Research* 43 (2009): 2209-2223.
[8]M. W. LeChevallier, N. J. Welch and D. B. Smith, "Full-scale studies of factors related to coliform regrowth in drinking water", *Applied and Environmental Microbiology* 62 (1996): 2201-2211.
[9]R. C. Jamieson, R. J. Gordon, K. E. Sharples, G. W. Stratton and A. Madani, "Movement and persistence of fecal bacteria in agricultural soils and subsurface drainage water: A review", *Canadian Biosystems Engineering* 44 (2002): 1.01-1.09.
[10]C. Ferguson, A. M. D. Husman, N. Altavilla, D. Deere and N. Ashbolt, "Fate and transport of surface water pathogens in watersheds", *Critical Reviews in Environmental Science and Technology* 33 (2003): 299-361.
[11]M. L. Anderson, J. E. Whitlock and V. J. Harwood, "Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments", *Applied and Environmental Microbiology* 71 (2005): 3041-3048.
[12]L. D. Ogden, D. R. Fenlon, A. J. Vinten and D. Lewis, "The fate of Escherichia coli O157 in soil and its potential to contaminate drinking water", *Int J Food Microbiol* 66 (2001): 111-7.
[13]G. A. Olyphant, J. Thomas, R. L. Whitman and D. Harper, "Characterization and statistical modeling of bacterial (Escherichia coli) outflows from watersheds that discharge into southern Lake Michigan", *Environmental Monitoring and Assessment* 81 (2003): 289-300.
[14]A. M. Sadeghi and J. G. Arnold, "A SWAT/microbial sub-model for predicting pathogen loadings in surface and groundwater at watershed and basin scales", in, *Total Maximum Daily Load* (St Joseph, 2002), 56-63.
[15]M. V. Yates, "Classical indicators in the 21st century - Far and beyond the coliform", *Water Environment Research* 79 (2007): 279-286.
[16]R. Collins and K. Rutherford, "Modelling bacterial water quality in streams draining pastoral land", *Water Research* 38 (2004): 700-712.
[17]F. C. Curriero, J. A. Patz, J. B. Rose and S. Lele, "The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948-1994", *American Journal of Public Health* 91 (2001): 1194-1199.
[18]K. R. Reddy, R. Khaleel and M. R. Overcash, "Behavior and transport of microbial pathogens and indicator organisms in soils treated with organic wastes", *Journal of Environmental Quality* 10 (1981): 255-266.
[19]N. Schaffter and A. Parriaux, "Pathogenic-bacterial water contamination in mountainous catchments", *Water Research* 36 (2002): 131-139.
[20]N. R. Khatiwada and C. Polprasert, "Kinetics of fecal coliform removal in constructed wetland", *Water Science and Technology* 40 (1999): 109-116.
[21]A. Schafer, P. Ustohal, H. Harms, F. Stauffer, T. Dracos and A. J. B. Zehnder, "Transport of bacteria in unsaturated porous media", *Journal of Contaminant Hydrology* 33 (1998): 149-169.
[22]T. K. Stevik, G. Ausland, J. F. Hanssen and P. D. Jenssen, "The influence of physical and chemical factors on the transport of E-coli through biological filters for wastewater purification", *Water Research* 33 (1999): 3701-3706.
[23]C. M. Ferguson, B. G. Coote, N. J. Ashbolt and I. M. Stevenson, "Relationships between indicators, pathogens and water quality in an estuarine system", *Water Research* 30 (1996): 2045-2054.
[24]Emma Britton, Simon Hales, Kamalesh Venugopal and Michael G. Baker, "The impact of climate variability and change on cryptosporidiosis and giardiasis rates in New Zealand", *J Water Health* 8 (2010): 561-71.
[25]Thomas B. Atherholt, Mark W. Lechevallier, William D. Norton and Jeffrey S. Rosen, "Effect of rainfall on Giardia and crypto", *American Water Works Association Journal* 90 (1998): 66-80.
[26]B. Goldar and N. Banerjee, "Impact of informal regulation of pollution on water quality in rivers in India", *Journal of Environmental Management* 73 (2004): 117-130.
[27]L. Egwari and O. O. Aboaba, "Environmental impact on the bacteriological quality of domestic water supplies in Lagos, Nigeria", *Revista De Saude Publica* 36 (2002): 513-520.
[28]T. A. Stukel, E. R. Greenberg, B. J. Dain, F. C. Reed and N. J. Jacobs, "A longitudinal-study of rainfall and coliform contamination in small community drinking-water supplies", *Environmental Science & Technology* 24 (1990): 571-575.
[29]J. S. Hansen and J. E. Ongerth, "Effects of time and watershed characteristics on the concentration of cryptosporidium oocysts in river water", *Applied and Environmental Microbiology* 57 (1991): 2790-2795.
[30]C. Mons, A. Dumetre, S. Gosselin, C. Galliot and L. Moulin, "Monitoring of Cryptosporidium and Giardia river contamination in Paris area", *Water Research* 43 (2009): 211-217.
[31]S. R. Crane, P. W. Westerman and M. R. Overcash, "Die-off of fecal indicator organisms following land application of poultry manure", *Journal of Environmental Quality* 9 (1980): 531-537.

[32]J. A. Entry, R. K. Hubbard, J. E. Thies and J. J. Fuhrmann, "The influence of vegetation in riparian filterstrips on coliform bacteria: II. Survival in soils", *Journal of Environmental Quality* 29 (2000): 1215-1224.
[33]M. A. Faust, "Relationship between land-use practices and fecal bacteria in soils", *Journal of Environmental Quality* 11 (1982): 141-146.
[34]H. J. Kibbey, C. Hagedorn and E. L. McCoy, "Use of fecal streptococci as indicators of pollution in soil", *Applied and Environmental Microbiology* 35 (1978): 711-717.
[35]D. N. Mubiru, M. S. Coyne and J. H. Grove, "Mortality of Escherichia coli O157 : H7 in two soils with different physical and chemical properties", *Journal of Environmental Quality* 29 (2000): 1821-1825.
[36]S. Pedley and G. Howard, "The public health implications of microbiological contamination of groundwater", *Quarterly Journal of Engineering Geology* 30 (1997): 179-188.
[37]J. B. Robertson and S. C. Edberg, "Natural protection of spring and well drinking water against surface microbial contamination .1. Hydrogeological parameters", *Critical Reviews in Microbiology* 23 (1997): 143-178.
[38]R. L. Tate, "Cultural and environmental-factors affecting longevity of escherichia-coli in histosols", *Applied and Environmental Microbiology* 35 (1978): 925-929.
[39]S. Godfrey, F. Timo and M. Smith, "Relationship between rainfall and microbiological contamination of shallow groundwater in Northern Mozambique", *Water Sa* 31 (2005): 609-614.
[40]C. Hagedorn, D. T. Hansen and G. H. Simonson, "Survival and movement of fecal indicator bacteria in soil under conditions of saturated flow", *Journal of Environmental Quality* 7 (1978): 55-59.
[41]R. A. E. Barrell and M. G. M. Rowland, "Relationship between rainfall and well water-pollution in a west-african (gambian) village", *Journal of Hygiene* 83 (1979): 143-150.
[42]M. H. Barrett, K. Johal, G. Howard, S. Pedley and M. Nalubega, "Sources of faecal contamination of shallow groundwaters in Kampala", in O. Sililo, ed., *Groundwater: Past Achievements and Future Challenges* (Leiden, 2000), 691-696.
[43]J. S. Jean, H. R. Guo, S. H. Chen, C. C. Liu, W. T. Chang, Y. J. Yang and M. C. Huang, "The association between rainfall rate and occurrence of an enterovirus epidemic due to a contaminated well", *Journal of Applied Microbiology* 101 (2006): 1224-1231.
[44]D. Kay, J. Watkins, C. A. Francis, A. P. Wyn-Jones, C. M. Stapleton, L. Fewtrell, M. D. Wyer and D. Drury, "The microbiological quality of seven large commercial private water supplies in the United Kingdom", *Journal of Water and Health* 5 (2007): 523-538.
[45]T. Kistemann, T. Classen, C. Koch, F. Dangendorf, R. Fischeder, J. Gebel, V. Vacata and M. Exner, "Microbial load of drinking water reservoir tributaries during extreme rainfall and runoff", *Applied and Environmental Microbiology* 68 (2002): 2188-2197.
[46]G. Nichols, C. Lane, N. Asgari, N. Q. Verlander and A. Charlett, "Rainfall and outbreaks of drinking water related disease and in England and Wales", *Journal of Water and Health* 7 (2009): 1-8.
[47]M. K. Thomas, D. F. Charron, D. Waltner-Toews, C. Schuster, A. R. Maarouf and J. D. Holt, "A role of high impact weather events in waterborne disease outbreaks in Canada, 1975-2001", *International Journal of Environmental Health Research* 16 (2006): 167-180.
[48]H. Auld, D. MacIver and J. Klaassen, "Heavy rainfall and waterborne disease outbreaks: The Walkerton example", *Journal of Toxicology and Environmental Health-Part a-Current Issues* 67 (2004): 1879-1887.
[49]L. A. Hicks, C. E. Rose, B. S. Fields, M. L. Drees, J. P. Engel, P. R. Jenkins, B. S. Rouse, D. Blythe, A. P. Khalifah, D. R. Feikin and C. G. Whitney, "Increased rainfall is associated with increased risk for legionellosis", *Epidemiology and Infection* 135 (2007): 811-817.
[50]M. Hashizume, B. Armstrong, S. Hajat, Y. Wagatsuma, A. S. G. Faruque, T. Hayashi and D. A. Sack, "Association between climate variability and hospital visits for non-cholera diarrhoea in Bangladesh: effects and vulnerable groups", *International Journal of Epidemiology* 36 (2007): 1030-1037.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

A NEW 'ORIENTALISM': THE QUEER 'OTHER' ON THE FINAL FRONTIER

Teresa JEWELL

Independent Researcher, Ontario, Canada
tess@thequeery.com

Keywords: Science fiction, Queer theory, Sexuality, Aliens, Others, Television

ABSTRACT

In its narrative construction, science fiction creates a space in which humanity and society can be examined critically from a safe distance by virtue of the time not being now, the place not here, the people not us. Particularly difficult questions of morality are often explored through alien species: sexual diversity and queer identities are two such subjects. Critics have seen science fiction's potential to introduce minority groups into the mainstream, and the occasional inclusion of queer characters and alternative lifestyles in science fiction television programmes has seemed to affirm this potential. However, instead of promoting sexual minorities, this narrative structure actually reinforces their exclusion from mainstream humanity by showing them to be alien, 'other'. This paper examines how the critical safety net of science fiction thus upholds the dominant position of heteronormative society in three recent series: *Farscape* (1999-2003), *Battlestar Galactica* (2004-2009), and *Stargate Universe* (2009-2010). It demonstrates the key lesson that while science fiction can promote inclusion and diversity, it must be approached carefully to avoid Othering the groups it claims to include.

INTRODUCTION

In describing his work on the original *Star Trek*, writer, director, and producer John Meredyth Lucas notes the power of science fiction to convey contemporary issues and controversial subjects: '*Set the story in outer space, in the future, and all of a sudden you can get away with just about anything, because you're protected by the argument that, "Hey, we're not talking about the problems of today, we're dealing with a mythical time and place in the future"*'.[1] This '*cognitive estrangement*',[2] to use Darko Suvin's term, is created by situating science fiction in a different time and/or space, even though its narrative elements are drawn from present-day society; the use of aliens to explore questions of morality and human identity is another such example. Given that the genre has been used to examine race relations and promote the inclusion of visible minorities, as in the *Star Trek* franchise,[3] it is surprising to note the relative paucity of sexual minorities in science fiction television programs. Although the number has increased since Wendy Gay Pearson's assertion in 1999 that '*sf shows seem to be the last hold-outs in a medium that is rapidly accommodating itself to the idea that there really are lesbian and gay people in the "real" world that television claims, however peculiarly, to reflect*',[4] even when alternative lifestyles and queer characters are depicted, their frequent association with aliens can actually reinforce their exclusion from mainstream humanity by showing them to be alien, 'other'. When these aliens are further presented in the context of manipulation, betrayal, and physical or mental harm, as they often are in the series to be discussed, the messages regarding sexual diversity are even less affirmative. This paper examines the relationship between queer theory and the alien Other in science fiction before looking at how three recent North American series, *Farscape* (1999-2003), *Battlestar Galactica* (2004-2009), and *Stargate Universe* (2009-2010), can be seen to uphold the dominant position of heteronormative society despite their inclusion of sexual minorities. While science fiction can promote inclusion and diversity, progress must still be made in casting queer characters as Self and not Other.

ISBN 978-0-9566139-1-2

Science fiction, Orientalism, and queer theory

In order to investigate the portrayal of sexual diversity and queer characters in recent science fiction television programs, it is first necessary to examine the use of the Other in science fiction and its intersections with queer theory. Although it may at first seem strange to associate science fiction television with Edward Saïd's discussion of 'Orientalism as a Western style for dominating, restructuring, and having authority over the Orient'.[5] However, the two are related in more than just their mutual use of the Other, as Arab in the second case and alien in the first. As Ziauddin Sardar and Sean Cubitt explain, the roots of science fiction are within Western culture and its Orientalist mode of thought: '*science fiction employs the particular constellations of Western thought and history and projects those Western perspectives on a pan-galactic scale. Science fiction re-inscribes Earth history, as experienced and understood by the West, across space and time*'.[6] This relationship is not entirely negative: while science fiction, like Orientalism, demonstrates '*a certain* will *or* intention *to understand, in some cases to control, manipulate, even to incorporate, what is a manifestly different (or alternative and novel) world*',[7] its use of outer space also provides '*a site at which liberatory practices may (or may not) occur*'.[8] In other words, science fiction has the potential to promote alternatives to dominant Western culture just as much as it can subordinate them. *Farscape*, for example, explores both possibilities in the relations between its white, North American, heterosexual male hero and his entourage of exotic aliens.

Within this framework, the figure of the alien Other can be used either to demonstrate what is not human or to explore less mainstream aspects of humanity by veiling them behind the protective shield of cognitive estrangement, and it is in this context that science fiction becomes important to queer theory. As Wendy Gay Pearson explains,

> *If we take as the central task of queer theory the work of imagining a world in which all lives are livable, we understand queer theory as being both utopian and science fictional, in the sense of imagining a future that opens out, rather than forecloses, possibilities for becoming real, for mattering in the world. We begin to understand the importance of sf as a genre for exploring these very possibilities, as well as for interrogating the consequences of societies and futures in which conditions render the lives of many unbearable.*[9]

Considering the minority status of non-traditional sexual activity and identity in both science fiction and Western culture as a whole, the position of the alien can be seen as analogous to that of the queer: '*To be different, or alien, is a significant if familiar cultural metaphor which marks the boundaries and limits of social identity. It allows difference to be marginalised and any dissonance to be smoothed away, thus confirming the dominance of the centre over the margins*'.[10] Associating culturally non-standard sexualities with aliens in science fiction thus reifies the position of the queer as Other within even future society. From the perspective of the queer audience, this discourse of difference, alienation and subordination also further encourages identification with the Other through the common experience of marginalization; even when alternative sexualities are not explicitly present, this can leave a series open to 'wholesale reclamation' in which any and all characters can be read allegorically as expressing queer desires and identities.[11]

As we will see in the following sections, culturally non-standard sexuality is thus kept at the margins even as it is included in *Farscape*, *Battlestar Galactica*, and *Stargate Universe*. In keeping with Patricia Meltzer's description of queer theory as '*conceptualiz[ing] desire in ways that destabilize the naturalized correlation between sex, gender, and sexuality and make room for shifting identities within the categories of gender and sexuality*',[12] for the purposes of this discussion I will be using the term 'queer' rather than 'gay' in order to include a wider diversity of identities and sexual practices such as fetishes, triads, bisexuality, and homosexuality. In order to consider a broad spectrum of examples, I will also only be considering select episodes from each series to give a general idea of all three.

DISCUSSION

'Testin' for reactions to the freak show': Dangerous diversity in Farscape

In *Farscape* (1999-2003), astronaut John Crichton is accidentally catapulted into a distant part of the universe and stranded aboard a biomechanoid ship carrying a host of escaped alien prisoners. Onboard Moya, Crichton's '*able, middle-class, heterosexual white male*'[13] values bump up against the more colourful bodies and cultures of various outcast aliens: Pa'u Zotoh Zhaan, a blue, plant-evolved priestess; Ka D'Argo, a tentacled warrior reminiscent of

ISBN 978-0-9566139-1-2

Star Trek's Klingons; Dominar Rygel the XVI, a lascivious '*two-foot green slug*'[14] and former ruler of over 600 billion subjects; Chiana, a monochromatic nymphomaniac; and Aeryn Sun, an initially emotionless soldier from a militaristic race, to name only a few. Since Crichton is the only human in this part of the universe, all questions of sexual or cultural difference must be explored through aliens—even Aeryn's human-looking Sebacean species demonstrates little racial diversity.[15] The show's direct exploration of alternative sexuality often occurs through Chiana, whose '*questionable (and generally self-serving) morals*'[16] show no limits in her desire to seek out new experiences and types of pleasure; however, there are also queer potentialities in the friendship between Crichton and D'Argo and in the antagonistic power play between Crichton and his nemesis, Scorpius. In the first case, the homosexual tensions serve to introduce humour into a scene, whereas with the leather bondage gear-clad Scorpius the purpose is usually to heighten the sense of physical and psychological danger. Overall, despite its frequent presentation of alternative lifestyles and values, *Farscape* can be seen to represent sexual diversity as threatening. Since Jes Battis does an excellent job of investigating both the humorous and sinister elements of the male pairings in several episodes, I will begin by focusing in detail on a key episode he barely mentions: '*Won't Get Fooled Again*' (2.15, first aired August 18, 2000).

In this episode, Crichton has been captured by an enemy race and is undergoing psychological torture consisting of hallucinations calculated to manipulate his emotions and sense of reality. In the imaginary world constructed around him, figures from Moya are recast in human roles and mixed with familiar people and places on Earth in increasingly nonsensical ways. Crichton isn't fooled by the illusion, but, as a helpful vision of Scorpius eventually informs him, deceiving him isn't the point:

> *Now that the Scarrans know I'm after you, they want to know why. This is their, um, standard method of interrogation. Induced delusions to break down all mental defences. The Scarrans aren't trying to fool you, John. They're trying to break you. This is a deliberate attempt to drive you into insanity and it's working. Your mind is about to crack.*[17]

There have been certain sexual overtones to the events so far, mostly in terms of an unusual degree of promiscuity among Crichton's re-envisioned friends: D'Argo has become a partying womanizer, and he and Aeryn's character begin embracing passionately upon meeting each other for the first time. In a later encounter at the same bar, their affection expands to include Chiana in the form of a floozy named Jessica, who fawns over Fairchild before turning her attention to Crichton himself.

After Scorpius's revelation, however, the sexual aspect of the delusions reaches an overwhelming intensity as their logic deteriorates exponentially. Crichton next wakes up to the most shockingly discordant scene yet. He is still fully clothed but strapped to a medical examining table with his legs in the air, surrounded by his female shipmates dressed in classic fetish wear: the blue priestess-turned-psychiatrist Zhaan is clad in black latex with elbow-length gloves; Chiana sports a leather Japanese schoolgirl outfit; and Aeryn has traded in her doctor's lab coat for a PVC nurse uniform complete with invasive medical probes that she prepares for use. The three women fight for Crichton's interest, responding to his protests by saying this is 'nothing you haven't already fantasized about', but their advances are cut short by Rygel, the slug-like alien now bearing a leather bondage mask and a whip that he uses to begin beating Crichton. The first three fetishes may have been disturbing to Crichton, but now that a male (and the annoying Rygel, no less) is participating, Crichton is pushed past his limit and manages to escape.

This scene alone might seem the apex of absurdity and sexual threat; however, the delusions descend into further chaos as Crichton breaks down. After surviving a fall from a high window, Crichton is accosted by a male police officer (Peacekeeper Captain Bialar Crais, his nemesis from Season One) in red ladies' pumps who attempts to arrest him and proposes a sex hotline for his 'one phone call': '*I recommend Trixie: 976-Triple 5-LOVE*'. In the next delusion, which takes place in a romantic spot overlooking the night-time city lights, a comically flamboyant and lisping D'Argo confesses to Crichton that he has been '*thinking about you in a very different way*' and is '*wondering if you would mind participating with me in a little Luxan "bonding ritual"*.' Crichton emits an agonized '*Oh no*' as D'Argo strokes his cheek and rejects his protestation before shifting into the next—and arguably most disturbing—act. This time, Crichton awakens in a psychiatrist's office to find his mother dressed in a silk robe and negligee, asking him to explain the Oedipus Complex to her in incestuously physical terms. She first drapes a stockinged knee over his shoulder where he sits on the couch, then attempts to draw him into an embrace that he only narrowly resists. At this point, John is nearly broken.

ISBN 978-0-9566139-1-2

The progression from one scene to the next as Crichton descends toward madness is thus as follows: BDSM (bondage and sadomasochism) fetishes activity involving women, to fetish activity with men, to crossdressing, to homosexuality, to incest. One could argue that these correspond to increasingly primitive instincts being unlocked in Crichton as his intellectual defences are bypassed; however, we know that the Scarran torturer is in direct control over the delusions through his ability to change the scene and make characters ask specific questions (as he does when a 'real' Aeryn rescues Crichton but refuses to let him leave until he tells her about Scorpius, breaking the illusion). Aside from the Oedipus Complex vignette, the scene on the medical table draws the clearest links between humour and terror, sexual 'deviance' and insanity, in what is, in effect, an attempted rape on both a mental and physical level.[18] The absurdity of seeing aliens in human fetish costumes is juxtaposed against their complete domination over the restrained Crichton and his unwillingness to participate. Even though they are waiting for him to decide between them, they are still not taking 'no' for an answer. Also, the women and Rygel are filmed from Crichton's viewpoint, so that the audience is invited to experience the insanity—the hilarity, the fear—just as Crichton is experiencing it. When Rygel begins the whipping, we too are being whipped. The threat of sexual violation is the central element, made explicit without the humorous element when Crichton is forced to confront the advances of his sexualized mother.

I have lingered on '*Won't Get Fooled Again*' because it contains multiple examples of how *Farscape* shows alternative sexual practices to be a source of threat; however, as Battis notes, '*almost every episode of* Farscape *features body politics and bodily relations of some sort*'.[19] Another key episode that actually does combine race with sexuality and threat is '*Twice Shy*', in which Chiana purchases a stereotypically beautiful black slave girl named Talikaa in order to rescue her from abusive owners. The slave seems to capture the 'pansexual' Chiana's imagination almost immediately, although compared with the latter's assertive behaviour toward men, her tentative seduction and kiss seem '*to be a moment of rather innocent sexual expression on Chiana's part,…an exploratory gesture*'.[20] The embrace is cut short by Aeryn,[21] and Chiana uncharacteristically shifts her sexual attention to her, then to Crichton and even Scorpius. But this level of libido is '*excessive, even for [Chiana]*', as Aeryn observes while spurning her advances, and Talikaa is quickly named the cause of behavioral changes among the entire crew. Unfortunately, Talikaa is revealed to be '*a homicidal spider-alien who wants to devour everyone on the ship*',[22] her touch '*infect[ing her] victims and stimulat[ing] exaggerated neural functioning*' that she then harvests.[23] This is what has happened to Chiana: her sexual response to the female slave is not purely a product of her own desire but rather of the effects of the toxin in her system. The audience has been given cues to the danger that Talikaa presents right from the beginning from the traders' desire to be rid of her and from the distorted perspective we see through her eyes as she looks at her new prey; we know she is more than her appearance suggests. The seduction scene with Chiana is correspondingly eerie as we watch her alternate between entranced and disconcerted through Talikaa's eyes. Talikaa, then, presents us with the figure of both a racial and sexual minority in terms of her blackness and bisexuality; however, these traits are inseparable from and thus intrinsically linked with the threat she poses to the crew. Interestingly, she is not just a danger to their lives: Chiana is equally horrified to discover that Talikaa has robbed her of 'strongest trait', the libidinous potential she initially evinced.[24] It is only by killing the slave and destroying her nest that Chiana's sex drive—and by extrapolation the heteronormative society onboard Moya—is restored.

These two episodes taken together with the extensive homosocial and homoerotic tensions explored through the male aliens on *Farscape* thus paint a nuanced picture of sexual diversity: even though queer potentialities are not generally treated by Crichton or the alien characters as taboo or transgressive, their use to convey humour and threat underlines a discomfort with behaviour outside of the heteronormative standard. Thus, while *Farscape* is more imaginative than earlier (and even subsequent) science fiction television programs in its depictions of sexuality, an analysis of the portrayals of these practices reveals them to still be subordinated to the dominant heterosexuality explored through the challenging but ultimately enduring relationship between Aeryn and Crichton.

'Transcend[ing] the barriers that separate people': Bisexuality and betrayal in Battlestar Galactica

The reimagined *Battlestar Galacitca* series (2004-2009) differs drastically from *Farscape* in that its '*"ragtag fleet" of Colonial survivors on their quest for Earth*'[25] is exclusively human, the 'others' in this case being not aliens but robotic beings originally created by the humans themselves. These Cylons are '*intelligent robots…[that] were used as slaves and soldiers to fight humanity's wars*' before they rebelled, eventually destroying the Twelve Colonies and attempting to

ISBN 978-0-9566139-1-2

exterminate the entire human race in an ongoing genocidal war.[26] Aside from the fact that each Cylon is a copy of one of only seven possible bodies, they are physically indistinguishable from humans at even the cellular level.[27] Thus, barring a few key exceptions, the Cylons are understandably perceived by the Colonials as the ultimate nemesis due to both their hatred of humans and ability to masquerade perfectly as them. Given this extreme distrust of the Cylon 'race', it is interesting that two out of three queer relationships shown in *Battlestar Galactica* involve at least one Cylon and have strong overtones of manipulation and betrayal.

The first of these relationships occurs in Season Three as a brief triad between two female Cylons, Caprica Six and D'Anna, and Dr. Gaius Baltar, a human of questionable loyalty who hopes to discover he is a Cylon in order to absolve himself of guilt and become a hero to a new people after aiding the Cylons in their initial genocide and subsequent acts of subjugation against the human colonies.[28] Their three-way romantic and sexual relationship occurs off screen, but the audience does get to see the context in which it begins and ends. From Baltar's perspective the relationship seems to be about survival and manipulation: when the Cylons suspect him of involvement with an anti-Cylon virus and D'Anna tortures him for information in '*A Measure of Salvation*', Baltar follows the advice of an internal vision of Six to '*find the holes in her psyche*' in order to make her stop, finally looking directly at D'Anna as he confesses love for the Six making love to him in his mind: '*You have to believe in me; you're all I have left…I believe in you. I love you. I love you with all my heart*'.[29] D'Anna is fooled by the deception and is next seen sharing a bed with Caprica Six and Baltar, although this is the only image of intimacy shown between them. It is interesting to note that even the colours in this episode serve to create a strong visual link between torture, manipulation, and sex in this episode. In an otherwise utilitarian white and silver Cylon environment, red is the colour of the chaise on which Baltar awakens before learning he is about to be tortured, the internal Six's dress in his vision, the blood that stains his white shirt during the torture, and finally the sumptuous red bedsheets he shares with D'Anna and Caprica Six. The bed is especially conspicuous, standing alone in the middle of an empty white room.

Despite her manipulation by Baltar, D'Anna also comes to seem invested in the relationship purely for selfish reasons once she begins to believe that Baltar can help her in her quest to see the faces of the 'final five' Cylons. In '*The Eye of Jupiter*', Caprica Six expresses unhappiness with the progress of their union, feeling marginalized by the stronger heterosexual bond that has formed between Baltar and D'Anna: '*I'd like to think we three have shared something, have transcended the barriers that separate people and yet, somehow, the more time goes on the more I find myself on the outside looking in*'.[30] While Baltar initially makes a characteristic attempt to smooth over this shift in the focus of the relationship, D'Anna cuts him off to announce coldly that they are '*finished*' with her, and her consolation that they still love her despite their desertion comes off as empty and emotionless. The result is that *Battlestar Galactica* shows a heterosexual pairing '*transcending*'—to use Baltar's term—the non-traditional triad, its exploration of an alternative sexual lifestyle through a manipulative relationship between Cylon and pro-Cylon Others still ultimately serving to reify the norm.

Two other occurrences of queer relationships appear in *Battlestar Galactica*, though neither is given any visibility during a regular episode of the show. The made-for-television movie '*Razor*' situated between seasons Three and Four tells the story of Admiral Helena Cain, who discovers that her lesbian lover Gina Inviere is actually a Cylon, once again using the Cylon threat to portray alternatives to heterosexuality. Their relationship is initially shown to be loving and respectful, an unusual exception to Cain's general disregard for the feelings of others and '*morally monstrous*' orders that kill families and cripple civilian ships.[31] However, even this temporary happiness is clouded by the fact that the audience has already been introduced to Gina as an abused Cylon prisoner in Season Two, so that we know in advance how catastrophically the relationship is doomed to end.[32] Still, despite Gina's betrayal as a fully-aware Cylon agent carrying out her orders, her feelings for Cain are never proven to be a mere act; instead, when her identity is uncovered, she hesitates when contemplating shooting Cain and appears to show regret at the necessity of her actions. Cain, by contrast, reveals herself to be the true monster in the degree to which she enacts her revenge. Othering Gina to the point of an 'it', Cain authorizes the use of torture and gang rape in her interrogation: '*Since it's so adept at mimicking human feeling, I'm assuming that its software is vulnerable to them as well, so...Pain, yes, of course. Degradation. Fear. Shame. I want you to really test its limits. Be as creative as you need to be*'.[33] Overall, the heterosexual norm is protected by the flashback nature of '*Razor*' in that the relationship has ended and its members been killed even before the movie begins.[34] Also, by presenting Cain as an inhuman, monstrous figure in her treatment of Gina and even her own crew, *Battlestar Galactica* makes her even less sympathetic than

ISBN 978-0-9566139-1-2

the Cylon enemy. Thus, this portrayal can be seen to warn against lesbian relationships as a source of mutual betrayal and violation.

There is one queer relationship within *Battlestar Galactica* that is not couched in such negative overtones, but its status is marginalized within the context of the series to the extent that there is almost nothing to discuss. In a set of 10 3- to 5-minute Internet 'webisodes' entitled '*The Face of the Enemy*', Lieutenant Felix Gaeta is revealed to be bisexual when he shares a quick kiss with his partner Lieutenant Louis Hoshi, although this is the full extent of the sexual closeness shown between them.[35] In fact, most of the episode focuses on a previous relationship Gaeta had with a treacherous female Cylon: even the kiss between him and this Number Eight is given more screen time, more intimacy than the kiss with Hoshi. What limited representation there is is positive, however. Hoshi's commitment to finding Gaeta's disappeared ship is juxtaposed against the increasingly murderous acts and revelations of past betrayal by Number Eight, although it should be noted that Eight is the one who paradoxically saves Gaeta while Hoshi return to Galactica in defeat. Overall, in their limited scope, the webisodes do indeed show homosexuality between human males as a superior alternative to further manipulation and betrayal by the Cylon Other.

Given these three relationships, while *Battlestar Galactica* depicts a variety of sexual diversity among its recurring cast, its association of non-traditional relationships with the manipulative and treacherous Cylons or monstrous humans like Admiral Cain still serves to undercut the inclusive, sympathetic potential of these explorations. Furthermore, when a homosexual relationship is shown as being nurturing and supportive, these scenes are restricted to online webisodes or television movies outside the regular series. Considering the large amount of heterosexual sex depicted in detail throughout *Battlestar Galactica*, then, the consistently implied and problematized nature of the queer relationships reveals how much the former remain privileged and the latter marginalized.

'You can do better than that': Lesbian lead on Stargate Universe

Even more so than *Farscape* and *Battlestar Galactica*, *Stargate Universe* (2009) '*boldly go[es] where no science fiction television has gone before*' by casting one of its regular roles as a lesbian.[36] Most importantly, this show provides a significantly more positive portrayal of a queer relationship between two human women—both also visible minorities—although the context is not entirely unproblematic. By setting itself in the present and maintaining an all-human cast, for one, the series limits itself in its ability to use cognitive displacement to explore human sexuality, and its focus in the lesbian relationship tends to be on domestic relations more than on sexuality.

In an overall premise similar to that of *Star Trek: Voyager*, a mixed crew of military and civilians is accidentally stranded on a starship 'several billion light-years from home'.[37] Unfortunately, the starship, named *Destiny*, is travelling on autopilot on a course headed away from Earth, and its systems are so damaged after millennia of disuse that the crew is unable to use the stargate to return home. Communication with Earth is still possible, however; this is effectuated only through alien devices simply called '*communication stones*' that allow the user to '*physically take control of an individual at the other end*'.[38] In other words, a person onboard the ship can trade consciousness with a person on Earth also using a stone, each inhabiting the other's body for a time.

The lead character in question is Camile Wray (Ming-Na), the leading member of the civilian International Oversight Committee (IOA)[39] on *Destiny* and one of the three major figures vying for control of the ship.[40] Her long-time lover Sharon (Reiko Aylesworth) remains on Earth. In an official interview on the MGM Stargate website, actress Ming-Na discusses how this relationship was envisioned by the show's executive producers:

> *What was lovely was that both Robert [C. Cooper] and Brad [Wright] wanted Sharon and Camile's relationship to be the most steady and the most fulfilling to experience, I guess, for the audience, and they've been together for 12 years and there's a lot of real love and companionship…and I found that so refreshing and so lovely and that it just wasn't about 'chicks in space'. I was really really happy that they were going to be portrayed with such dignity and love and respect.*[41]

Compared with the various heterosexual relationships portrayed on the show, that of Camile and Sharon is indeed the most nurturing and positive. For example, while the audience is shown a jealous Col. Everett Young accusing his wife of infidelity back on Earth while he himself has an extramarital affair with a medic onboard *Destiny*, we only see the lesbian relationship in scenes of domestic comfort, care, and support on Camile's visits

ISBN 978-0-9566139-1-2

home.[42] However, it is these visits—Camile's separation from Sharon as well as her need to occupy another's body in order to see her—that problematize their status within the context of the show.

First, Camile's is the only major romantic relationship separated by several billion light-years, if we discount Young's rocky marriage due to his affair on *Destiny*. Consequently, the heterosexual pairings depicted take precedence in most episodes: we are frequently shown conventional couples enjoying sexual intimacy including the developing closeness between Chloe Armstrong and First Lt. Matthew Scott, for example, whereas Sharon's existence is largely segregated from the primary action of the show due to her absence on Earth. While other lovers have sex in the episode '*Life*', Camile can only sketch a landscape to help her remain hopeful for the retirement she planned to share with her partner.[43] Camile's lesbianism is thus all but invisible in her daily life aboard *Destiny* despite having 'the most steady' relationship; indeed, her unwavering commitment to an absent lover also precludes all possibilities for exploring the character's queer potential except when she switches bodies and goes home. Seen from this perspective, Camile's homosexual lifestyle can be interpreted as having been left far behind.

In addition to the marginalized position Camile's sexual preference occupies onboard *Destiny*, her actual time spent with Sharon on Earth is complicated by the body switching. When Camile first appears on Sharon's doorstep in the body of an unfamiliar white woman, she must ask about a household 'ugly chair' in Mandarin Chinese to prove her identity.[44] Beyond this moment, Camile is only shown visually as the person within and not the host body on the outside, and the difference is largely ignored except for an occasional comment. As such, the potential issues this foreign body raises for the kissing and implied sexual relations that occur are also initially sidestepped: we see Camile—not the white woman—kissing Sharon, and Sharon treats Camile as if she were there in her own flesh. This may positively suggest that bodies are distinct from identity, following the premise of the *Star Trek: The Next Generation* episode '*The Host*' and equally ignoring the question of what is appropriate use of the borrowed body.[45] However, the switching also presents a barrier to their interactions not experienced by the heterosexual couples physically present together on *Destiny*. This barrier is made explicit in '*Sabotage*', in which their relationship becomes a vehicle for exploring disability when Camile switches into the body of a quadriplegic for a three-week sojourn on Earth. Her decision to sacrifice mobility and independence in this extremely Other body is proof of her love and dedication to Sharon, as is Sharon's eagerness to care for her; however, the episode's almost exclusive focus on the difficulties of living with a disability means that little to no time is spent on developing their back story or relationship. As such, even within an episode focusing specifically on their relationship, the queer element is marginalized.

CONCLUSION

Despite this criticism of queer relations within recent science fiction television programs, the fact remains that some progress has been made since *Star Trek*'s failure to portray its infamous '*infinite diversity in infinite combinations*'.[46] Series such as *Battlestar Galactica* and *Stargate Universe* have finally included queer figures among their casts of main characters, although only the latter has done so during a regular season on network television or given an unqualified portrayal of a positive queer relationship. To fit within the size constraints of this paper, I have only been able to provide an overview of the dominant discourses and elements at play within a few select episodes of each series; as such, each is open to future research in more detail. Further programs to consider could include *Caprica* (2009-present), a *Battlestar Galactica* prequel, and the British relaunched version of *Doctor Who* (2005-present) and its spin-off, *Torchwood* (2006-present), all of which involve depictions of queer relationships.

Even from this brief discussion, however, it emerges that science fiction programs remain heteronomative and suspicious of queer sexuality more often than not, choosing to maintain its association with Otherness via aliens and robotic beings, or even simply through their narrative structures. Short of creating a queer-centric program such as '*Queer as Space Folk*', this marginalization in traditional television may be inescapable. Still, despite its limitations, science fiction can be seen to reveal the universality of alienation, as Majel Barrett-Roddenberry once said: '*[This] kind of television program…will help ensure future generations will not have to carry the memory of persecution and trial and subjugation with them all their lives…with the message of course that we are all alien in one way or another*'.[47] From this perspective, the queer Other as alien is thus still ultimately a reflection of the self.

ISBN 978-0-9566139-1-2

REFERENCES

[1] Lucas, John Meredyth. Quoted in William Shatner with Chris Keski. *Star Trek Memories*. (1993) New York: HarperCollins, p. 243.
[2] Suvin, Darko, 'On the Poetics of the Science Fiction Genre,' *College English* 34 (1972). p. 372.
[3] The problematic use of aliens to represent race issues in the original *Star Trek* series has been explored in detail in Bernardi, Daniel Leonard. *Star Trek and History: Race-ing Toward a White Future*. (1998) New Brunswick: Rutgers University Press.
[4] Pearson, Wendy Gay. 'Alien Cryptographies: The View from Queer.' *Queer Universes: Sexualities in Science Fiction*. (2008) Ed. Wendy Gay Pearson et al. Liverpool: Liverpool University Press, p. 14.
[5] Saïd, Edward W. *Orientalism*. (1978) New York: Vintage Books, p 3.
[6] Sardar, Ziauddin and Sean Cubitt, eds. *Aliens R Us: The Other in Science Fiction Cinema*. (2002) Virginia: Pluto Press, p. 2.
[7] Saïd, Edward W. *Orientalism*. (1978) New York: Vintage Books, p 12. Italics in original.
[8] Casper, Monica J. and Lisa Jean Moore. 'Inscribing Bodies, Inscribing the Future: Gender, Sex, and Reproduction in Outer Space.' *Sociological Perspectives* 38.2 (Summer, 1995), p. 313.
[9] Pearson, Wendy Gay et al., eds. *Queer Universes: Sexualities in Science Fiction*. (2008) Liverpool: Liverpool University Press, p. 5.
[10] Wolmark, Jenny. *Aliens and Others: Science Fiction, Feminism and Postmodernism*. (1994) Iowa City: University of Iowa Press, p. 27.
[11] Tulloch, John, and Henry Jenkins. *Science Fiction Audiences: Watching* Dr. Who *and* Star Trek. (1995) New York: Routledge, p. 261.
[12] Melzer, Patricia. *Alien Constructions: Science Fiction and Feminist Thought*. (2006) Austin: University of Texas Press, p. 17.
[13] Battis, Jes. *Investigating Farscape: Uncharted Territories of Sex and Science Fiction*. (2007) New York: I.B. Tauris, p. 3.
[14] "Won't Get Fooled Again." *Farscape*. Dir. Rowan Woods. Sci-Fi Channel. Aug. 18, 2000.
[15] One exception is "Jeremiah Crichton," in which Crichton is stranded on an Earth-like planet populated by a primitive tribe of dark-skinned people outwardly identical to humans. However, this episode focuses on religion and superstition, and the only sexual possibilities shown are heterosexual ("Jeremiah Crichton." *Farscape*. Dir. Ian Watson. Nine Network. July 30, 1999.)
[16] Booker, M. Keith. *Science Fiction Television*. (2004) Westport, Connecticut: Praeger Publishers, p. 169.
[17] "Won't Get Fooled Again." *Farscape*. Dir. Rowan Woods. Sci-Fi Channel. Aug. 18, 2000.
[18] The Season Four episode "Prayer" explicitly refers to Scarran torture as 'mind-rape' when Aeryn is subjected to similar treatment, further cementing this analogy ("Prayer." *Farscape*. Dir. Peter Andrikidis. Sci-Fi Channel. February 17, 2003.)
[19] Battis, Jes. *Investigating Farscape: Uncharted Territories of Sex and Science Fiction*. (2007) New York: I.B. Tauris, p. 120.
[20] Ibid. p. 106.
[21] It is interesting to note that the BBC airing of this episode cuts the actual kiss so that Aeryn interrupts before Chiana and Talikaa are able to touch lips. However, the audience is still shown Talikaa touching Chiana's breast as Aeryn enters the room. Apparently fondling but not kissing was considered appropriate for viewing ("Twice Shy." *Farscape*. Dir. Kate Woods. BBC2. Jan. 13, 2003.)
[22] Battis, Jes. *Investigating Farscape: Uncharted Territories of Sex and Science Fiction*. (2007) New York: I.B. Tauris, p. 106.
[23] "Twice Shy." *Farscape*. Dir. Kate Woods. Sci-Fi Channel. Jan. 24, 2003.
[24] Ibid.
[25] Eberl, Jason T. ed. *Battlestar Galactica and Philosophy: Knowledge Here Begins Out There*. (2008) Malden, MA: Blackwell Publishing Inc., p. x.
[26] 'About.' Battlestar Galactica. *Syfy*. (No date) http://www.syfy.com/battlestar/about/, accessed Apr. 18, 2010.
[27] "Litmus." *Battlestar Galactica*. Dir. Rod Hardy. Syfy. Nov. 22, 2004. Television.
[28] Koepsell, David. 'Gaius Baltar and the Transhuman Temptation.' In *Battlestar Galactica and Philosophy: Knowledge Here Begins Out There*. (2008) ed. Jason T. Eberl. Malden, MA: Blackwell Publishing Inc., p. 248-49.
[29] "A Measure of Salvation." *Battlestar Galactica*. Dir. Bill Eagles. Syfy. Nov. 10, 2006. Television.
[30] "The Eye of Jupiter." *Battlestar Galactica*. Dir. Michael Rymer. Syfy. Dec. 15, 2006. Television.
[31] Jensen, Randall M. 'The Search for Starbuck: The Needs of the Many vs. the Few.' In *Battlestar Galactica and Philosophy: Knowledge Here Begins Out There*. (2008) ed. Jason T. Eberl. Malden, MA: Blackwell Publishing Inc., p. 108.
[32] "Pegasus." *Battlestar Galactica*. Dir. Michael Rymer. Syfy. Sept. 23, 2005. Television.
[33] "Razor." *Battlestar Galactica*. Dir. Félix Enríquez Alcalá. Syfy. Nov. 24, 2007. Television.
[34] Gina takes her own revenge and kills Admiral Cain in "Resurrection Ship: Part 2." *Battlestar Galactica*. Dir. Michael Rymer. Syfy. Jan 13, 2006. Television. She commits suicide before the end of Season Three.
[35] "The Face of the Enemy." *Battlestar Galactica*. 10 webisodes. Syfy. Dec. 12, 2008 to Jan. 12, 2009. http://video.syfy.com/shows/battlestar/the_resistance/face-of-the-enemy--webisode-1/v1133102. Accessed May 22, 2010.
[36] Booker, M. Keith. *Science Fiction Television*. (2004) Westport, Connecticut: Praeger Publishers, p. 169.
[37] "Air: Part 1 and Part 2." *Stargate Universe*. Dir. Peter DeLuise. Syfy. Oct. 2, 2009. Television.

ISBN 978-0-9566139-1-2

[38] Ibid.

[39] Fans of the first two television series, *Stargate: SG-1* and *Stargate: Atlantis*, may have an initial bias against Camile due to her status as a member of the IOA. In this series, the International Oversight Advisory often interfered negatively with the Stargate Program and was portrayed as an adversary rather than an ally (for example, see "The Ties that Bind." *Stargate: SG-1*. Dir. William Waring. Sci-Fi Channel. Aug. 5, 2009. Television.) *Stargate Universe* quickly problematizes both the Stargate personnel and archaeologist Dr Nicholas Rush so that none of the three sides appears more sympathetic than the other; however, the initial casting of the lesbian character as an IOA member is interesting in this light.

[40] 'Main Character: Ming-Na as Camile Wray.' *Official Stargate Website*. http://stargate.mgm.com/view/character/170/index.html, accessed Apr. 29, 2010.

[41] Ming-Na. 'Sharon in my Life.' Interview. *Official Stargate Website*. (Nov. 23, 2009) http://stargate.mgm.com/view/content/1843/index.html, accessed Apr. 22, 2010.

[42] "Sabotage." *Stargate Universe*. Dir. Peter DeLuise. SyFy. May 7, 2010. Television.

[43] "Life." *Stargate Universe*. Dir. Alex Chapple. Syfy. Nov. 11, 2009. Television.

[44] Ibid.

[45] Roberts, Robin. *Sexual Generations: "Star Trek: The Next Generation" and Gender*. (1999) Chicago: University of Illinois Press, p. 115. Up until this point in *Stargate Universe*, however, the show has not yet allowed body switching between persons of different genders, precluding all opportunities for exploring questions of gender identity.

[46] Joyrich, Lynne. 'Feminist Enterprise? Star Trek: The Next Generation and the Occupation of Femininity.' *Cinema Journal* 35 (1996), p. 68.

[47] Barrett-Roddenberry, Majel. Quoted in Kozinets, Robert V. 'Utopian Enterprise: Articulating the Meanings of Star Trek's Culture of Consumption.' *The Journal of Consumer Research* 28.1 (Jun., 2001), p. 75.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

THE TIMES THEY ARE NOT A CHANGING: NEOLIBERAL INTERPELLATION, REVISITING THE FRANKFURT SCHOOL

Rodolfo, LEYVA

PhD Student, Kings College London, London, UK
rodolfo.leyva@kcl.uk.ac

Keywords: Media/Cultural Studies, Political-Economy, Social-Psychology

ABSTRACT

There is now a growing critical literature documenting the rise of media driven consumer cultures within contemporary liberal-democratic societies like those of the US and the UK.[1] According to this literature, there is a reciprocal relationship between the mainstream media and state-corporate economic interests, where the corporate media and ubiquitous corporate advertisements spread a consumer ideology that seeks to interpellate young people into market discourses and identities that are congruent with a neoliberal teleology.[2] This article seeks to add to this literature by reexamining and building on the social-psychological and political-economic works of the classic Frankfurt School theorists namely; Theodor Adorno, Max Horkheimer, and Herbert Marcuse, to analyze the hegemonic cultural-ideological dimensions of modern neoliberal societies. Referencing mostly US and UK empirical and textual evidence, with a brief section on related international case studies, this article also argues that the societal adoption and promotion of what can be considered neoliberal norms and values (e.g., consumerism, materialism, self-interestedness, and competitiveness) can lead to detrimental ecological, psychological, and societal effects that have potential global consequences.

INTRODUCTION

Over the last the thirty years the institutional and ideological political-economic frameworks of contemporary UK and US societies have been shaped and underpinned by neoliberalism- a political-economic philosophy that calls for the implementation of free-market logic and practices into all forms of civic, public, and private life.[3] Formulated by an elite group of mostly Western intellectuals and business leaders, neoliberalism is framed under a narrow ontological view of human nature; one that argues that human beings are mostly rational actors whose predominantly self-interested and calculating behaviors are best channeled and most beneficial to society when operated in a completely free-market arena.[4] Given such ontological assumptions, advocates of neoliberalism argue that in order to achieve national and international peace, harmony, prosperity, and parliamentarian or republican forms of democracy[5], governments should completely deregulate businesses, sell off the public sector to private interests, and facilitate international free-trade.[6] The UK and US governmental implementation of neoliberal ideas and practices during and after the administrations of Thatcher and Reagan, has given rise to the dominance of multi-national corporations- which are increasingly taking control of the major societal institutions that govern and shape the lives of UK, US, and global populations (i.e. educational, political, economic, and cultural institutions).[4]

In this article, I will focus on the multifaceted ways in which neoliberalism shapes and influences contemporary UK and US popular culture. Throughout this paper I will make the case that essentialist neoliberal ideas and behaviors such as extreme forms of individualism, competitiveness, and self-interestedness, are congruent and interwoven with the norms and values of conspicuous consumption, meritocracy, and materialism that

ISBN 978-0-9566139-1-2

characterize the contemporary media driven consumer cultures of the UK and US.[1] Furthermore, in disseminating these interlocking norms and values via monopolized, conglomerated, and multinational media and culture industries, UK and US corporate and government representatives (who either own, manage, or use these industries) are in effect carrying on an institutionalized and global hegemonic project, a hegemonic project that seeks to maintain, sustain, and reproduce, the contemporary neoliberal political-economic order[7] - as well as divert criticism, attention, or alternative thought from it[8] - and whose teleological ends aim to create a global society of rational, individualistic, and mostly uncritical consumers.[9] To build this argument, I will first review and appropriate elements from the theories of the classic Frankfurt school, (which apply a unique blend of sociology, psychology, and political-economy to the study of modern capitalist culture). Secondly, I will then reformulate and merge those elements with Louis Althusser's concept of interpellation (explained later in this piece) to analyze excerpts from contemporary UK and US popular culture, with specific references made to youth culture as contemporary youth are exposed to an unprecedented number of insidious and ubiquitous corporate adds. Third, I will review some related international case studies that offer some indirect empirical evidence from which some of the effects of neoliberal ideas and practices can be extrapolated. Lastly, by examining excerpts from UK and US popular culture and reviewing related international case studies I highlight how neoliberal consumer culture co-opts youth culture and point to some of the potential detrimental effects that can occur when aspects of neoliberal ideology are internalized by individuals, but will acknowledge that further research is needed to investigate whether, how far and in what ways the cultural experiences and values of contemporary individual youth are infused or inflected by neoliberal values. By no means is this an exhaustive and systematic analysis of popular culture, but I hope that it will spur debate nonetheless.

REVISITING THE FRANKFURT SCHOOL: THE RELEVANCE OF DEAD GERMANS

During the 1930's a group of exiled German sociologists, psychologists, philosophers, and literary scholars collectively known as the Frankfurt School, fled to the United States. Disheartened by what they saw as the totalitarian nature of both German and US societies, members of Frankfurt School sought to give an explanation as to why the working classes of the industrialized West failed to among other things, instigate a proletariat revolution. By combining the psychological insights of Sigmund Freud with the historical-materialist perspective of Karl Marx, the Frankfurt School developed 'critical theory'- a broad interdisciplinary cultural materialist approach that analyzes how macro-power structures shape and mediate the cultural practices and consciousness of individuals. Whilst seemingly outdated, classic critical theory according to media scholar Robert Babe[10] and political philosopher Dana Villa[11] continues to have an enduring relevance and explanatory/descriptive prowess, one that I will argue has specific elements that are especially useful in explaining the hegemonic processes of neoliberalism. What follows is an overview of two aspects of critical theory namely Horkheimer and Adorno's 'culture industry' theorizations, and Marcuse's 'repressive desublimation'.

In 1944, Horkheimer and Adorno published their seminal piece Dialectic of Enlightenment. In the chapter titled '*The Culture Industry: Enlightenment as Mass Deception*', they argue that post-war US capitalist culture has in essence created a techno-bureaucratic Weberian 'iron cage' that traps individuals (particularly those of the middle and working-classes) into a perpetual cycle of alienating work and consumption. US popular culture they argue, in the form of television shows, films, fashion, literature, art, and music, is for the most part industrialized, standardized, and commoditized state-corporate propaganda that promotes consumer capitalism and societal conformity, while simultaneously distracting the public from the source of their presumed economic hardships and alienating work. They open the chapter by arguing:

> *The sociological theory that the loss of the support of objectively established religion, the dissolution of the last remnants of pre-capitalism, together with technological and social differentiation or specialization, have led to cultural chaos is disproved every day; for culture now impresses the same stamp on everything. Films, radio and magazines make up a system which is uniform as a whole and in every part. Even the aesthetic activities of political opposites are one in their enthusiastic obedience to the rhythm of the iron system.*[12]

While at times extremely hyperbolic, Horkheimer and Adorno continue with the chapter by meticulously describing how market logic creates systemic rules that shape cultural values, artifacts, and aesthetics to the needs of capitalism, thereby turning critical citizens into intellectually passive consumers. The chapter offers elaborate and dense content analyses of how culture industries attempt this – however I will only discuss the two major insights that resonate throughout their arguments. First, starting from the classic Marxist idea that ideological

ISBN 978-0-9566139-1-2

obfuscations are rooted in the material structures of capitalist production,[11] Horkheimer and Adorno argue that culture and media industries/corporations, like the film, music, and television industries, disseminate an array of ideological constructs that serve to reinforce the capitalist order/agenda and erode alternative political-economic possibilities. The formulaic Hollywood films, generic pop-music, and banal television shows they argue, are saturated with affirming US state-capitalist ideals of rugged individualism, financial success, meritocracy, and red-baiting, Audiences of these texts are therefore left with a rather narrow, distorted, corporatized, and conformist reality – where one should not resist or challenge the political-economic order since there are equal opportunities for all to prosper from and within. Villa argues that anyone that thinks that Horkheimer and Adorno exaggerate this claim, need only look at contemporary Western movies and television shows which continue to pronounce the explicit American Dream thesis – that with enough hard work, persistence, and a little luck, anyone can achieve financial success. Villa continues [sic]: *'with one ideological catchphrase-endlessly recycled in TV and movie dramatizations of individuals who 'overcome the odds'-the grounding myth of society is established. An entire landscape of structural inequality and injustices is banished from out horizon'.*[13]

Secondly, Horkheimer and Adorno explicate the structural-institutional schemata that drive corporate behavior. Using Weber's concept of instrumental rationality, they argue that market logic drives these corporations to behave in rationalist ways that require constant economic growth. That is, the corporate structure manifests a metaphysical level of agency to ensure its survival amongst competing interests, and to guide the individual behavior of the people that run it. In citing the insatiable growth and monopolistic logic of the corporate structure, Horkheimer and Adorno, also predate sociologist C Wright Mill's *Power Elite,*[14] and media scholar's Herman and Chomsky's *'Propaganda Model',*[8] by arguing that as media-cultural corporations grow, they become interlocked with other private and state institutions all of which have structurally driven interests in spreading messages that endorse, legitimate, and promote elite state-corporate interests. They write: *'The dependence of the most powerful broadcasting company on the electrical industry, or of the motion picture industry on the banks, is characteristic of the whole sphere, whose individual branches are themselves economically interwoven.'*[15] In pointing to these interlocking directorates and systemic imperatives, Horkheimer and Adorno draw attention to both the vested interests involved in the production, packaging, and distribution of media-cultural texts, as well as the monopolistic drives of corporate structures. This is a particularly important observation, as the monopolization of media-cultural industries, which continues to this day, further narrows the ideological messages that are now spread to global audiences, while squeezing out alternatives- and in fact disseminating very explicit messages that state that there are no alternatives (I will come back to this point later in this piece).

Furthermore, in addition to describing the structural imperatives that force corporations to disseminate a specific consumer ideology that is in tune with their economic interests, Horkheimer and Adorno also analyzed the psychological dimensions that media and culture industries utilized to target individuals at the unconscious level. This aspect of critical theory is elaborated more by Marcuse who draws from Freudian psychology, and argues that media and culture industries prey on the individual's libidinal psychological drives.[16] By continuously inciting desire, corporations of all industries use media-cultural industries to manufacture wants and needs, and try to socialize individuals into consumptive modes that trap them in perpetual cycles of arousal, desire, and frustration. Marcuse referred to the process behind these trappings as 'repressive desublimation' – a social-psychological phenomenon where individuals are socially encouraged to give into their unconscious repressed desires, but only through socially constructive consumerist practices, i.e., through the consumption and fetishization of commodities (like pornography, sports cars, violent video games, or fast food), that once purchased and used, fail to redeliver satisfaction and gratification, requiring that individuals constantly consume more items to continue to fulfill their repressed needs. According to Marcuse, the state-corporate nexus requires that individuals are actively and constantly consuming, behaviors that require socialization via a total corporate environment- in which political-economic alternatives are dismissed, vilified, or omitted from public consciousness, and where creativity, libidinal urges, and individuality are to be channeled, or desublimated, through the constant consumption of goods and services that advanced industrial capitalism has made relatively affordable for everyone.

Lastly, whatever critical culture does develop like the environmentalist, feminist, and civil rights movements, Marcuse argues that the inclusive and rationalist nature of the media-cultural industries makes them co-opt even counter and sub-cultures (as they have profitable potential), strip them of their revolutionary potential, and sell it back to the public in sanitized or banal forms. For instance he notes that the cultural heroes, poets, rebels, and

ISBN 978-0-9566139-1-2

outcasts of 19th century literature and art that once represented an alternative to the status quo, have been transformed in advanced industrial society to be merely figures of entertainment that in some form or another uphold the status quo. According to Marcuse, if not totally resisted, challenged, and refused, the continuing corporate standardization and sanitation of culture, coupled with the spreading of repressive desublimation via insidious and psychologically honed advertisements, will lead to a one-dimensional society of alienating work and vapid consumption where consumer ideology becomes so ingrained into the public and becomes so hegemonic, that aptitude for critical resistance to it becomes relatively non-existent or ineffective in changing or seriously challenging the status-quo. And so, instead of creating a state of genuine freedom, advanced capitalism and consumer ideology are potentially more effective form of totalitarian social control. One that is largely self-imposed and more reminiscent of Huxley's Brave New World than Orwell's 1984 as individuals in some form or another become complicit in their own domination. According to Marcuse this form of totalitarianism can create an environment where:

> *Liberty can be made into a powerful instrument of domination. The range of choices open to the individual is not the decisive factor in determining the degree of human freedom, but what can be chosen and what is chosen by the individual. Free election of masters does not abolish the masters or the slaves. Free choice among a variety of goods and services does not signify freedom if these goods and services sustain social controls over a life of alienation. And the spontaneous reproduction of superimposed needs by the individual, does not establish autonomy, it only testifies to the efficacy of the controls.*[17]

ON NEOLIBERAL INTERPELLATION

Thus far I have described two major elements of classic critical theory: the culture industry theorizations and repressive desublimation. These elements help to illuminate the way in which media-cultural corporations operate on a structural, ideological, and psychological level to inculcate the public with a false consciousness. By distracting the public (with a variety of mostly manufactured/false needs and mundane entertainment), from the source of their economic hardships, the elite capitalist ruling class maintains power and domination. To analyze the modern neoliberal culture of the UK and US, (aspects of which are being adopted by cultures all over the world), I will start from these basic premises of classic critical theory and add the following arguments and theorizations. In disseminating a consumer ideology, contemporary media-cultural oligopolies, monopolies, and conglomerates also implicitly and explicitly disseminate a neoliberal ideology that stresses competitiveness and self-interestedness as ideal norms and values, and free-market capitalism and republican forms of democracy as the only viable political-economic arrangement. Thus, media-cultural corporations help to carry on the hegemony of neoliberalism by circumventing criticism from it on two levels. At the first level, they seek to completely inculcate audiences with an ideology that stresses that there is nothing more to life than shopping, individual upward mobility, and where political participation is limited to voting representatives into power. As Chomsky argues:

> *The people in the public relations industry aren't there for the fun of it. They're doing work. They're trying to instill the right values. In fact, they have a conception of what democracy ought to be: It ought to be a system in which the specialized class is trained to work in the service of the masters, the people who own the society. The rest of the population ought to be deprived of any form of organization, because organization just causes trouble. They ought to be sitting alone in front of the TV and having drilled into their heads the message, which says, the only value in life is to have more commodities or live like that rich middle class family you're watching and to have nice values like harmony and Americanism. That's all there is in life.*[18]

At the second level, if the first level of inculcation is not achieved and individuals become critical of the established order, in constantly promoting that there is no alternative to the established political-economic order and in reducing political discourse and images to sound-bytes, catchphrases, vacuous slogans and personalities, media-cultural corporations help to stymie the political-economic imagination of the public. This can create a sort of political dissonance, where even individuals who are critical of the status quo, do little to challenge it, as they are presented with few viable alternatives. And thus their inaction also serves to sustain the system by not directly challenging it. This is not to suggest that political-economic alternatives are non-existent, or that Francis Fukuyama is correct in famously declaring the end of the history.[19] Millions of individual activists and organizations both in Western and non-Western countries, continue to, and in some cases in the face of outright violent state-corporate repression, actively struggle against neoliberal corporate hegemony.[9] Klein[20] and Graeber[21] point to several anarchist and anti-globalization groups all over the world that are not only fiercely

ISBN 978-0-9566139-1-2

anti-neoliberalism, but that are also dedicated practitioners of alternative political-economic systems based on altruism, generosity, cooperation, and direct/participatory forms of democracy. The point I argue however, is even if one does not agree with these alternatives, their erosion from mainstream media (e.g., major newspaper and tv news shows), as demonstrated by Herman and Chomsky's 'propaganda model', and which is fully evident by turning on the news and seeing these movements either ridiculed, trivialized, or omitted all together, has a potentially debilitating effect on our political imagination and abilities to conceive of an alternative to the dominant neoliberal order, which continues to destroy the environment, exploit labor, and co-opt governments in an insatiable pursuit of growth and profits.

Furthermore, at this juncture I will introduce Louis Althusser's concept of interpellation and Pierre Bourdieu's concept of 'habitus'. From a conceptual standpoint, both can be understood as socialization theories that attempt to explain specific aspects of social reproduction. Althusser's interpellation theory refers to how individuals become ideological subjects through their immersion into society.[22] Althusser argues that ideological state apparatuses like the family, the media, and the educational system, teach or interpellate individuals into the practices and ideas of those systems. An individual is said to be interpellated, when he/she enacts the thoughts and behaviors that their society expects. When individuals do not conform to the system, the repressive state apparatuses like the police and the military step in to ensure conformity. However, Althusser also notes that there are multiple breaks, contradictions and points of contestation between different ideological state apparatuses and repressive state apparatuses that leaves room for agency and resistance. For example, individuals that refuse to join repressive state forces because they come from religious and pacifist backgrounds, or conscientious soldiers that refuse to take up arms, signify a clear contradiction between the interests of the state and the individual ideology of people who refuse to enact or enforce them. Interpellation works as a totalizing hegemonic process, but it is always contested by the complexities and variances of multiple social systems and sub-systems that teach and expose individuals to any number of different and in some cases conflicting ideologies.

Bourdieu's 'habitus' refers to an inherited set of cultural practices and dispositions possessed by individuals.[23] It is a term used to describe the sum of an individual's dispositions, tastes, habits, and acquired schemes of thought, perception and actions that are learned from interaction with autonomous structured social spaces like schools, courts and work. In occupying various social spaces, an individual internalizes any number of unique dispositions that enable him/her to learn, follow and modify the rules of those spaces. I contend that given the historic extension of neoliberal logic into seemingly all aspects of civic, public and private life, that contemporary youth living in neoliberal societies, are being exposed to a very narrow ideology and largely homogenized set of cultural practices and dispositions from which they can draw from, and which are aimed at permeating and shaping their individual habitus. By combining those elements of critical theory that I have described, i.e. the structural-economic imperatives and inclusive nature of media-cultural corporations, the erosion of political-economic alternatives, and the psychological apparatus of advertisements described by Marcuse's repressive desublimantion concept. And by merging them with Althusser's interpellation theory, what follows is a theoretical formulation that can be used to describe the hegemonic processes of contemporary neoliberal consumer culture, i.e. neoliberal interpellation. In the following segment, I will analyze excerpts from contemporary UK and US popular culture (e.g., music, television, fashion, and advertisements), and demonstrate how the insidious, ubiquitous and psychologically honed messages spread by media-cultural corporations seek to create a near total environment in which market logic and values become common sense, and which have the potential to shape the habitus of individual contemporary youth. However, I will acknowledge that empirical research needs to be conducted to investigate the micro-discourses and contexts of individuals living in these neoliberal cultures. What follows therefore, is merely a brief presentation of some of the dominant cultural-ideological messages that youth are exposed to; no claims will be made as to how youth interpret these messages.

GET RICH OR DIE TRYING

Before continuing, I should note that my analysis on the following pop-cultural examples is necessarily brief for page restrictions, and focuses on specific messages and media representations that I gathered from my participant observations and continued engagement with popular youth culture. As such, the following excerpts and observations were mostly chosen from my notes and memory, and is by no means an exhaustive list. Also, I will overlook other messages and representations that are disseminated by media, most notably gender and

ISBN 978-0-9566139-1-2

ethnic representations. This is mostly due to page restrictions, but also because neoliberalism operates as one size fits all ideology that attempts to intepellate everyone regardless of their ascribed statuses. For the most part, what seems to underpin most corporate media-cultural messages, as I will now show, is the implicit notion that anyone regardless of their gender, ethnic or class backgrounds, if they have the money for it, can enjoy the endless commodities and forms of entrainment being offered.

Among the more popular genres of music that according to record sales are most consumed by youth all over the world, are American-style hip hop and gangster-rap.[24] While these genres originally developed as working-class music that gave impoverished black youth an outlet to protest their conditions, it has since been co-opted by international media corporations for commercial gains. The pioneers of hip hop like Public Enemy openly questioned and challenged the status quo as evident in songs like *Fight The Power*. Commercial contemporary hip-hop is completely devoid of such explicitly anti status quo substance (echoing Marcuse's argument on the inclusive nature of media-cultural industries). Take for example, the lyrics from the following table.

Table 1: Excerpts from Popular Rap Songs

Jay Z's ***Money Cash Hoes***	**50 Cent's** ***In Da Club***	**Snoop Dog's** ***Drop it like it's hot***	**Estelle's** ***American Boy***
Flavors robust platinum and gold touch Y'all rap now, fast money lets slow it up Niggaz try to stop Jay-Z to no luck Roc-A-Fella foreva CEO what what Us the villains, fuck your feelings While yall playa hate we in the upper millions	And the plan is to put the rap game in a choke hold I'm feelin' focused man, my money on my mind I got a mill out the deal and I'm still on the grind Now shawty said she feeling my style, she feeling my flow Her girlfriend wanna get bi and they ready to go	Uh! I'm a nice dude, with some nice dreams See these ice cubes, see these Ice Creams? Eligible bachelor, million dollar boat That's whiter than what's spilling down your throat The Phantom, exterior like fish eggs The interior like suicide wrist red	Just another one champion sound Me and you about to get down Who the hottest in the world right now Just touched down in London town Bet they give me a pound Tell them put the money in my hand right now Tell the promoter we need more seats We just sold out all the floor seats

Source: Google Lyrics.

The songs displayed on the table, which can be heard in clubs from London to Los Angeles, have very explicit messages. These songs very much reaffirm the premises of capitalism and individual upward mobility. While Public Enemy's songs have clear community conscious messages, the lyrics from the more commercially successful rappers listed above celebrate opulent wealth, conspicuous consumption, individual success, and in many cases seek to incite envy and desire from their audiences. The music videos to these songs, for instance, have the rappers constantly sporting their 'bling', driving expensive cars or boats (which most viewers do not know are rented for the video shoot and not actually owned by the rappers), and all together taunting their viewers for not being as rich and as cool as they are. The consumers of these texts, which number in the millions, are therefore exposed to clear messages that promote materialism, consumption and individual upward mobility, all of which are congruent with a neoliberal ideology that promotes the same values. Other forms of popular music like rock and pop, while perhaps less explicit on notions of materialism, have lyrical content filled with cliché lyrics about love and romantic relationships. For instance as of May 24, 2010 when I began writing this piece, the Billboard and MTV UK Top 10 (which measures the relative popularity of songs and albums from the UK and the US) has on its list the songs: *Break Your Heart* by Taio Cruz, *Your Love is My Drug* by Kessha and Bob FT Bruno *Mars's Nothing On You*. While from different genres including R&B and Pop, the lyrics from these songs are essentially standardized and interchangeable as they all contain lyrics regarding love and romantic relationships, mirroring the state of popular music that Horkheimer and Adorno complained about over fifty

ISBN 978-0-9566139-1-2

years ago. Even though these songs are not explicitly enticing their audiences to buy Rolexes and BMW's, they are not exactly challenging the status quo either, which indirectly serves to sustain the system. When mainstream musicians and pop-stars do become openly political, it is usually in the form of supporting a particular candidate and thus upholding the representative democratic system. For example, in rapper Puff Daddy's '*vote or die*' campaign, Puff Daddy and a score of other celebrities encouraged youth to be political by voting for one of the major candidates during the 2004 US Presidential race. In the US, similar celebrity-led voting campaigns are carried on during major election cycles. Being political is therefore equated with a consumer choice method of voting someone into power; this is as far as a political imagination that these musicians promote and present to youth audiences.

Television is similar. Popular television shows, like the British *X-Factor* or the US counterpart *American Idol,* are essentially engineered to promote the myth of success where through hard work and dedication contestants can reach stardom and financial success (it should also be noted that contestants for these shows start out in the tens of thousands). In more subtle ways, these shows also promote cut-throat competition and individualism, as the purpose of the shows is not to bring together talented musicians and singers to collaborate and construct original music, but rather to crown an individual star who throughout most of the show's span will be singing old pop-songs. Other popular and internationally viewed television shows include *Friends, Sex In the City,* and *Top Gear*. To quickly sum up the content, *Friends* is an American sitcom about yuppies that spend their time in coffeehouses that largely resemble Starbuck's branches and talk about their careers and relationships. *Sex In The City* is a show about four independent upwardly mobile women who focus on liberal-feminist issues like careerism vs. starting a family, and who constantly fixate on shoes, wardrobe, and handbags; where those goods come from and under what conditions they are produced is never mentioned. Audiences are therefore presumably left to assume that they were produced with relatively little exploitation, or – since the show is interrupted by commercials for brand clothing – it is likely that corporate executives prefer that audience disregard such thoughts all together. *Top Gear* is a reality show in which men race expensive vehicles across different terrains and fetishes over gas-guzzling vehicles. The show incidentally fails to mention that the racing of those vehicles is needlessly contributing to global warming and ecological destruction, small as the scale may be. The messages from these tv-shows are as implicit and explicit as the songs I listed, and also present the receivers with values concerning individual upward mobility, conspicuous consumption, and materialism. If cultural artifacts and expressions like songs and tv shows are portraits that reflect society, a devoted consumer of popular UK and US culture can be left with a rather corporatized and distorted reality: a picture where everyone is in pursuit of money and fame, but where there are seemingly no real social problems, only individual struggles. A devoted consumer might also be forgiven for not knowing that both the UK and US are at war with Iraq and Afghanistan where hundreds of thousands of Iraqis and Afghanis have died as a direct result of UK and US bombings. Or for not knowing that in the US, the wealthiest nation on the planet, 17 million children go hungry each day,[25] since the overwhelming majority of cultural and media texts from UK and US societies that I examined simply do not concern themselves with these issues. What usually remains is therefore a distorted corporatized reality.

Western multinational fashion, sports and electronics industries epitomize this distorted corporatized reality that neoliberalism breeds. Far from the Western shopping malls and high-streets, clothes, soccer balls and electronics are made in third-world sweatshops and factories, where workers who are often teenagers and young adults, work in deplorable conditions for long hours and slave-wages.[26] Yet the ads (which can in their own right be considered a sort of post-modern Western cultural artifact that syncs art, technology, psychology, and economics), for these products, conveniently leave these facts out of the picture. Consumers are instead presented with rock stars, celebrities and otherwise attractive people sporting the latest brands, enticing youth to purchase them. Companies like Virgin, Apple, Nike and Starbucks argues Klein, no longer market products, but brand identities and lifestyles.[27] To be an Apple person means to be creative and artistic, or to be a Nike person is to be athletic, but under no circumstance is the individual consumer to be concerned with the environmental destruction and labor exploitation behind commodity production. Moreover, the same structural-economic imperatives that force corporations to grow and to externalize costs, also forces them to identify themselves as brands and not as product manufacturers. Many commercial industries, like fashion, sports and food corporations, operate under what economists refer to as monopolistic competition, a competitive market state in which multiple firms produce nearly identical products as that of their competitors, and thus individual corporations have to differentiate themselves through brands, labels and mass advertising. For example, H&M, Topshop, American Apparel and the Gap, all sell similar clothing for young demographics; what is different,

ISBN 978-0-9566139-1-2

however, are the labels and marketing approaches. In order to maintain their market share, these companies have to keep the costs of production low but also have to market ideas and identities, not products whose manufacturing is outsourced and contracted out to elaborate networks of second and third parties. Thus Starbucks, for example, does not sell coffee like Dunkin Donuts or Pret does, it sells community and ambiance. And so in this distorted and corporatized reality, exploitation never happens and individual consumers are free to negotiate between the advert texts and create a unique sense of identity by picking and choosing between mass produced goods and pre-packaged identities. Villa notes:

> *Mass culture provides us with innumerable ways of distinguishing ourselves from the cars we drive, to the films we like, to the fashion labels we prefer, to our vacation destinations all of which are no more than commodities aimed at specific markets or demographics. Even what counts as eccentric characteristics-the funny hat, the bow tie, the unorthodox facial hair, the sultry walk or the "in your face attitude"- is no more than a codified signifier cribbed from television or film.*[28]

This is not to suggest that popular culture is completely devoid of critical and even resistance messages; in fact probably more so now than in the mid 20th century that the Frankfurters wrote about. Contemporary examples include the films *The Corporation, Capitalism A Love Story* or the BBC's *What Happened To Our Dreams of Freedom* televised documentary series, all three of which give a chilling and critical look at neoliberal capitalism. Anarchist inspired bands like Rage Against the Machine, The Clash, R.E.M, and Radiohead have found huge commercial success. Books exposing the contradictions of American and British governments and written by leftist authors like Noam Chomsky and David Harvey can often be seen climbing the New York Times Bestsellers list. However, according to philosopher Rick Roderick, one of the main insights from the works of the classic Frankfurt School is that the Frankfurters argued that Western culture is not completely devoid of critical thought, but rather that critical thought becomes 'banalized.[29] The 'banalization' of culture, argues Roderick, happens when the culture industries take serious issues like poverty, war and racism, and trivializes them by turning them into mere forms of entertainment, meant to be consumed. For instance, anti-capitalist books by leftist intellectuals David Hill, Howard Zinn and even Karl Marx are sold at Barnes and Noble and Borders, and are meant to be read while sipping on Starbucks coffee. Music by socially conscious bands like Anti-Flag, the Beatles, and Bad Religion is featured on MTV and VH1[30] in between commercials for acne medicine, GAP clothing, and cell phones. The very word 'revolution' is co-opted for capitalist gains and can be heard on commercials for Nike and Apple. Revolutionaries like Che Guevara and Malcolm X, who sacrificed their lives to bring better living conditions for their respective impoverished peoples, are turned into t-shirts that can be bought at most malls. And so whatever critical messages are introduced to youth audiences, they are trivialized, and ideas of resistance and revolution are meant to be purchased and consumed, but not practiced. Some of our most powerful socio-political ideas like community, environmentalism, revolution and even democracy are now treated as commoditized brand content or corporatized signifiers, consequently cheapening them. Sub-cultures do often arise as disenchanted youth become resistant to dominant corporatized discourses and identities, such as the LA and London psychobilly Greasers who dress in 50's clothing, and listen to a non-commercial form of music that merges punk with rockabilly or the ravers from underground raver scenes that listen to electronic house music while popping ecstasy. But even these sub-cultures that are not completely co-opted by corporations are hardly a threat to the established neoliberal state-corporate order. At most they provide a space where participants can escape the gaze of the corporate lens, but is in no way a challenge to it. As sociologist Dick Hepdige writes:

> *Youth sub-cultural styles are meaningful mutations capable of embodying a symbolic refusal of the social consensus upon which Western democracies depend. But in the end, no amount of sub-cultural incantation can alter the oppressive mode in which the commodities used in a sub-culture have been produced.*[31]

Admittedly this has been a brief overview of contemporary popular culture, one perhaps too brief to make generalizations from, and one that reads more like an open and shut case, mirroring the totality of classic critical theory. However, what was presented was meant to highlight how macro structures like the state and private sector co-opt culture to achieve full hegemony; it is not meant to suggest that full hegemony has been achieved and that we are all, or most of us, cogs in the neoliberal machine. That is, the consumption of modern media, which is specifically designed to achieve neoliberal interpellation, does not necessarily mean that the consumers in question are shaped inexorably by such media. I personally, for example, very much enjoy Western popular culture like music and films. In fact the most enjoyable part of my day is when after spending hours reading dense social theory, economics and political philosophy texts, I can come home and watch episodes of *The Simpsons, Peep Show, The Boondocks, BBC News.* I do not then assume, of course, that the BBC news is telling me

ISBN 978-0-9566139-1-2

the whole truth or that Homer Simpson could ever be an astronaut. The 1930's hypodermic needle model of the media has been largely discredited. That said, two things need to be considered. The first is that contemporary corporate media and advertisement is unprecedented in scope, size and scientific development and in no way resembles the corporate advertising of the past. Western populations are now exposed to 1500-3000 scientifically honed corporate messages a day,[32] and as Marcuse wrote about forty years ago, corporations continue to research and enhance psychological marketing and publicity strategies to target individual consumers at the unconscious and subliminal level so as to incite desire and override their rationality in order to mold them into eternal and loyal consumers.[33] As Rowan reports, corporations are in hot pursuit of the holy grail of marketing, the buy button.[34] Researchers are using MRI machines (originally meant to scan for tumors and brain damage) to carve out the most objective ways to predict which logos and adverts will most trigger a subconscious response from consumers. Given that advertising and these types of research cost corporations billions of dollars, an Occam's Razor deduction would conclude that corporations would not spend billions on it if it did not work to satisfy their bottom line. In other words, modern advertising does not work like a magic bullet fired from the media gun into the consumer; it works more like a sawed-off shotgun, scatter shooting multiple messages in the direction of the consumer with hopes of hitting a target. Secondly, there is also a danger in completely ignoring the socializing effects of mass media and culture in favor of arguments for the fully active and sovereign consumer. Such notions, argue Babe and McGuigan, ignore the manipulative ideological intent of power-structures, and even mirror neoliberal notions of the fully rational agent.[35] If such primacy is given to the fully knowing fully active consumer, then for example there is no need for regulations on advertisements to children or for critical media education, since individual agents are seemingly invulnerable to manipulation. We are all pre-conditioned in some way by social structures that were in place before we were born. This is an ontological truism evident by the fact that we do not grow up in vacuums. Our identities and habits, as Bourdieu argues, come from a hotchpotch of sources and experiences that derive from our interaction with the social world and its structures. And so if neoliberal interpellation is to be resisted, it is also crucial to investigate and recognize the ways in which it manifests, conditions and manipulates our behavior.

SOME INTERNATIONAL CASE STUDIES

While little is known about how neoliberal ideology as I defined in the introduction actually influences youth populations, there are related media and psychological studies that show how neoliberal government and corporate practices (e.g. the globalization of corporate media and deregulated corporate marketing to youth) and congruent neoliberal values (self-interestedness, materialism, consumerism) have effected individuals and societies across the globe (the following studies contain much more depth but will be crudely reduced for purposes of page restrictions). In a 1998 study of Fijian social change, Becker concluded that after merely three years after the introduction of US media, the rates for eating disorders amongst Fijian adolescent girls went from 0 to 12%.[36] Other media studies offer similar testaments to influence of media on young peoples. For example, in a 2004 study by Story and French, the researchers found that there is a strong correlation between the food industry's intensive and increasing media marketing of fast-food, sweets and soft-drinks, and the rising rates of diabetes and obesity where 15% of youth are overweight, as compared to only 3% in 1980 when regulations on marketing to youth were stricter. In other neoliberal countries like the UK 45% of girls between ages 11 and 15 are overweight, in Australia 25% of children are overweight.[37] The findings for all these studies can be partly explained because, as Harvard child psychiatrist Susan Linn argues, children and teenagers are still developing cognitively, and despite all the arguments for the sovereign and active consumer, are not yet at the stage where they can fully resist an industry that spends billions of dollars a year to target them. Nor, she concludes, is it really fair for society and corporations to expect that level of agency from a vulnerable population that is still maturing.[38]

A few psychological studies have focused on how what can be considered congruent neoliberal values, e.g. self-interestedness, materialism, and competitiveness, have affected individuals that internalize them. In a 2007 article, psychologist Oliver James (using a multitude of survey data from 25 different countries) argues that countries like the US, UK, and Australia that have liberal-democratic, or what he referred to as selfish-capitalist, structures have the highest rates of mental illness among the countries studied. Other studies argue that individuals that focus on materialistic goals and motivations are more likely to suffer from anxiety, depression, substance abuse and personality disorders.[39] These studies however preliminary, seem to indicate that when

ISBN 978-0-9566139-1-2

direct or congruent aspects of neoliberal ideology are internalized by individuals, they can be psychologically detrimental. As Kasser et al. argue:

> *By using cross-cultural research on how values and goals are organized, we identified the aims most consistent with American corporate capitalism (ACC) (i.e., self-interest, financial success, and competition) and then reviewed a variety of literatures demonstrating how these aims conflict with and undermine pursuits long thought by psychologists to be essential to individual and collective well-being. These include helping the world be a better place, having committed, intimate relationships, and feeling worthy and autonomous. Further, we noted that ACC is built upon questionable assumptions about self-interest, competition, and the relationship between wealth and happiness.*[40]

END THOUGHTS

Throughout this chapter, I argued that there is a reciprocal and structural relationship between the media and corporate economic interests, where the corporate media and ubiquitous corporate advertisements spread a consumer ideology that seeks to interpellate young people into market discourses and identities that are congruent with a neoliberal ideology and teleology. The works of the classic Frankfurt school offered relevant insights helpful in demonstrating how market societies create their own logic about social order and cultural practices. By focusing on the political-economy of culture, my brief analysis of UK and US pop-culture examples and international media and psychological studies sought to bring to light issues of power, oppression and manipulation that are often lost in the mainstream of media and cultural studies. In sum, this paper is meant to highlight the dominant cultural-ideological messages that youth living in neoliberal societies are exposed to and how they are exposed to them, arguing throughout about the detrimental societal and psychological consequences of adopting such ideologies. Whether, how far and in what ways the cultural experiences and values of contemporary youth are infused or inflected by neoliberal values remains to be fully empirically explored.

REFERENCES

[1] Kenway, J. and Bullen, E. Consuming Children: Education-Entertainment-Advertising. (2001) Open University Press.
[2] Boyles, Deron. (ed). The Corporate Assault on Youth: Commercialism, Exploitation, and the End of Innocence. (2008) Peter Lang.
[3] Plehwe, D. Walpen, B. Neunhoffer, G. Neoliberal hegemony: a global critique. (2006) Routledge.
[4] Klein, N. The Shock Doctrine: The Rise of Disaster Capitalism. (2008) Picador.
[5] Note: Neoliberal ideology argues that representative forms of democracy are ideal political structures. This form of democracy should be contrasted to other forms of democratic thought which adhere to more inclusive and participatory forms such as direct democracy, demarchy, or consensus democracy. Advocates for the latter argue that representative forms of democracy are spurious manifestations of democratic thought that in practice yield power to a concentrated elite that use their political positions for self-interested gains. See for example, Graeber, D. *Fragments of An Anarchist Anthropology.* (2004) Chicago, IL: Prickly Paradigm Press, and Chomsky, N. *Media Control.* (2002) New York, NY: Seven Stories Press.
[6] Ellwood, W. No-Nonsense Guide to Globalization. (2002) New Internationalist/Verso.
[7] Harvey, D. *A Brief History of Neoliberalism.* (2005) Oxford: Oxford University Press.
[8] Herman, S. E. and Chomsky, N. *Manufacturing Consent: The Political Economy of the Mass Media.* (2002) New York: Pantheon.
[9] Patel, R. *The Value of Nothing: How to Reshape Market Society and Redefine Democracy.* (2010) Picador.
[10] Babe, R. *Cultural studies and political economy: toward a new integration.* (2009) Lexington.
[11] Villa, D. *Public Freedom.* (2008) Princeton University Press.
[12] Horkheimer, M., and Adorno, T. (1944/1993). *Dialectic of Enlightenment.* p. 1.
[13] Villa, D. *Public Freedom.* (2008) Princeton. p. 154-155.
[14] Mills, Wright. C. *The Power Elite.* (1956). Oxford University Press.
[15] Horkheimer, M., and Adorno, T. (1944/1993). *Dialectic of Enlightenment.* p. 2.
[16] Marcuse, H. *One Dimensional Man.* (1993/1964) Beacon Press.
[17] Marcuse, H. *One Dimensional Man.* (1993/1964) Beacon Press. p. 7-8.
[18] Chomsky, N. *Media Control.* (2002) New York, NY: Seven Stories Press. p. 22.
[19] Fukuyama, F. *The End of History and the Last Man.* (1992). Free Press.
[20] Klein, N. *No Logo: Taking Aim at the Brand Bullies.* (2000) Canada: Alfred A. Knopf.

ISBN 978-0-9566139-1-2

[21] Graeber, D. *Fragments of An Anarchist Anthropology.* (2004) Chicago, IL: Prickly Paradigm Press.
[22] Althusser, L. *Ideology and Ideological State Apparatuses. Notes Toward and Investigation.* (2001). in Lenin and Philosophy and Other Essays. New York, Monthly Review Press.
[23] See section on Bourdieu in Gewirtz, S. and Cribb, A. *Understanding Education: A Sociological Perspective.* (2009) Polity Press
[24] Macphearson, A. *Dead wrong on hip-hop - the genre that won't die.* (2009). December 3, The Guardian.
[25] November 19, 2009, Democracy Now. http://www.democracynow.org/2009/11/19/as_wall_street_posts_record_profits
[26] Klein, N. *No Logo: Taking Aim at the Brand Bullies.* (2000) Canada: Alfred A. Knopf.
[27] Klein, N. *No Logo: Taking Aim at the Brand Bullies.* (2000) Canada: Alfred A. Knopf.
[28] Villa, D. *Public Freedom.* (2008) Princeton University. p. 149.
[29] http://www.youtube.com/watch?v=ixqH-j32aAE&feature=related
[30] Note: Popular cable-television networks designed for adolescents and young adults, and owned by Viacom a massive media conglomerate
[31] Hebdige, D. *Subculture: The Meaning of Style (New Accents).* (1981) Routledge. p. 84.
[32] See Fogel, G. truth or consequences: An analysis of public service announcements in contemporary society. (2006). Film and Media Studies Program, University of Rochester. vol 5 (1&2). pp. 95-102. Also
note that these statistics are representative of American sample sizes, although they are likely to be similarly representative for inhabitants of other Western nations and cities.
[33] See. Linn, S. *Consuming Kids: Protecting Our Children from the Onslaught of Marketing and Advertising.* (2004) Blackwell. And Crisp, R. *Persuasive advertising, autonomy, and the creation of desire.* (2004) Journal of Business Ethics. Vol 6 (5). Pp. 413-418.
[34] Rowan, D. *Neuromarketing: The search for the brain's 'buy' button.* (2008). June 29. Wired Magazine UK Edition
[35] Babe, R. *Cultural studies and political economy: toward a new integration.* (2009). Lexington. And McGugigan, J. *Critical Renewal?* (2003). Sociology. v. 37. pp. 591-597.
[36] Becker, A. *Television, Disordered Eating, And Young Women In Fiji: Negotiating Body Image And Identity During Rapid Social Change.* (2004). Culture, Medicine and Psychiatry v. 28. pp. 533–559.
[37] Beder, S. Varney, W. and Godsen, R. *This Little Kiddy Went to the Market: The corporate capture of childhood.* (2009) UNSW Press.
[38] Linn, S. *Consuming Kids: Protecting Our Children from the Onslaught of Marketing and Advertising.* (2004) Blackwell.
[39] Kasser, T. *The High Price of Materialism.* (2002). Cambridge-MIT Press.
[40] Kasser, T. Cohn, S. Kanner, A. D. and Ryan, R. M. *Some Costs of American Corporate Capitalism: A Psychological Exploration of Value and Goal Conflicts.* (2007). Psychological Inquiry. vol. 18, pp. 1–22. p. 18.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

UNDERSTANDING THE USER: THE MARGINALIZED OPEN SOURCE CONTRIBUTOR

Mekki MACAULAY

PhD Student, York University, Toronto, Ontario, Canada
mekki@osstrategy.org

Keywords: Open source, User, Developer, OSS, Consumer marketing, OpenOffice.org, Microsoft office

ABSTRACT

The level of adoption of open source software is a key metric of the health of open source ecosystems. Many open source projects struggle with adoption because they have a developer-oriented focus. The users of the product are marginalized and their contribution ignored. This paper examines the motivations of participation in open source ecosystems to explain the views on the role of the user. Then, using an interdisciplinary approach, the perspectives of the field of consumer marketing are compared to the current beliefs in the technical community to provide insights on the user perspective on software adoption. The case of user adoption barriers for the open source office suite OpenOffice.org, a competitor to Microsoft's Office suite, is considered using an empirical study.

The results of the study show that six factors affect user adoption behaviour, namely comfort with Microsoft, habituation to the Microsoft Office interface, dependence on the Microsoft Office file formats, user innovativeness, loyalty to Microsoft, and likelihood to have searched for alternatives to Microsoft Office. These factors are a marked contrast from the espoused beliefs in the technical and trade literature, but are accounted for in the extant consumer marketing literature, showing the benefit of using an interdisciplinary approach to gain insights on challenges faced in technical fields. The insights for the open source community are that concessions may need to be made in the design decisions for OpenOffice.org, such as changing the user interface, and default file format settings, in order to improve user comfort with the product; and, that targeted marketing for adoption of the product should focus on the type of user that is more innovative and less loyal to Microsoft.

INTRODUCTION

Open source software development is an approach to software development that recognizes the value of freely distributing the source code of the software to allow modifications and improvements to be made and redistributed. It is contrasted from closed source software development, which uses copyright law to prevent the modification and distribution of its source.

Open source software developers are typically volunteers who contribute their time to improve a product that is useful to them. This seemingly altruistic development model has been well researched. For example Bonaccorsi and Rossi found that many developers feel part of a unique culture and community, and work on an open source software project to foster a sense of belonging in the community.[1] Henkel found that developers contribute to an open source project as a means of getting support from the community for the aspects of the project that they are interested in and use most often.[2] Lerner and Tirole found that developing for an open source project gave programmers a venue to showcase their talent in the world, building their reputation, which sometimes led to paid work elsewhere.[3]

ISBN 978-0-9566139-1-2

Traditionally, there was a clear division between the participants in an open source project. The roles were clearly delineated as developers and users. Developers were the ones who shouldered all the work. They contributed source code to the projects. They debugged problems. They debated the merits of new features and added the as needed. They managed the projects and their evolution. Developers were the key participants who set the direction of a particular project and were the primary actors responsible for its success or failure. Open source projects were by developers, for developers.

Users, on the other hand, were thought to be passive participants who merely used the software. The developer community regarded them as 'leachers' – users who only took what was available for free and made no contributions back to the project or community. They were often spoken of in disdain by the 'real' contributors as opportunistic or problematic. Users were thought to have nothing to offer to open source projects. They were just freeloaders benefiting from the open source licensing terms. Snyder described the debate over the perception that those who don't contribute back in the form of code to an open source project were considered '*open source vampires*'.[4]

But this perspective only makes sense if you assume that roles are clearly delineated as developer and user. The reality of open source projects is quite different. Scipion described numerous other roles for participants in an open source ecosystem including new user support, collaboration facilitator, know-how sharer, evangelist, trainer, event organizer, donor, and user.[5] From the perspective of code contributions, these participants are all passive and these contributions to the open source ecosystem are not measured by the traditional scales. Yet they are clearly doing something important to contribute to the health of the open source ecosystem. Why are these roles marginalized?

Brian Proffitt, the community manager of the Linux Foundation clearly believes that users have a role to play in every open source project:

> *I reject the notion that any user is a freeloader or a leech. At the very least they are vectors for your software, getting it out there in real-world environments to show to other potential users. They'll see it in action, ask about it, and then perhaps pick it up for themselves. Maybe they'll buy it, contribute to it, or maybe they will just use it for free too. [They] will bend it, twist it, mash it, smash it, and shove it onto platforms and into tasks it was never designed to do, [a]nd when they break it, they will [h]and you an opportunity to make your software a little better. Use it.*[6]

Yet, there has only been limited research on motivations of user participation in open source communities. The primary reason is simply to satisfy a need. A user needs to get a job done and finds an open source product that can help. Another primary reason that users choose open source software is to save money, as they are frequently willing to trade off the warranty and support services offered by proprietary software companies in exchange for the inexpensive open source alternative.[7] This extant research shows that users clearly have a stake in open source software, but it is clear that their perspective is different from that of developers.

A better understanding of this different user perspective could lead to advances in software development. Pirillo described the opportunity thus:

> *What would the world of software be like if the inmates were running the asylum? I'd argue a lot more useful, and a lot more beautiful. But users are usually in the back seat when it comes to the evolution of a utility – from beginning to end. Let me put it to you this way: software is useless if there isn't anybody using it. The world of software is getting larger by the day, and more people are finding new and different ways to improve lives with digital code. Programmers suffer from a miscalculation of a user's wants, needs, and desires.*[8]

The following sections take a closer look at a specific example of the misunderstanding of users' wants, needs, and desires for open source software in the case of OpenOffice.org, an open source competitor to Microsoft's ubiquitous office suite. Using a research methodology based on consumer marketing principles, a study was conducted that examined the product-specific and user-specific characteristics that lead to non-adoption of OpenOffice.org. The insights from this research shed light on some of the perception differences between developers and users, and suggest means of bridging the gap.

ISBN 978-0-9566139-1-2

METHODOLOGY

The purpose of this experiment was to gather data on why users failed to adopt OpenOffice.org as an alternative to Microsoft Office. Assessing the opinions of users directly presented several challenges. Users are not always aware of their own decision-making. Their responses to questions may show significant bias, and may be filled with an extensive 'laundry list' of issues with no clear prioritization of the factors that actually lead to a change in behaviour. Colgate and Lang addressed this issue in consumer marketing research by designing a methodology that examines users who have considered switching software products, but didn't.[9] This approach has many benefits in that it looks back at a past decision instead of speculating on a potential future decision. It also has higher reliability and highlights the factors that are immediately impeding adoption—the factors that lead to the greatest confusion and misunderstanding—such that they can be addressed directly.

In order to identify the product factors in contention, an extensive review of the technical and trade press was conducted, and an exhaustive list of suggested switching barriers and motivators was developed. The following table (Table 1) describes these product factors.

Table 1: Switching barriers and motivators from technical and trade press

	Select factors theorized to affect product adoption
Switching barriers	Fewer capabilities / features; Compatibility problems with MS Office; User resistance to change; User preconceptions; Comfort with Microsoft; Lack of training and reference material options for OO.org; Exclusivist culture of developers and users of OO.org; Users used to MS Office interface; Slower performance of OO.org; Security flaws in OO.org; Less powerful dictionary & grammar check in OO.org; Discounted versions of MS Office available; OO.org is less flexible than MS Office
Switching motivators	Novel / more features; Social pressure; Bias against Microsoft; OO.org is available at no cost; Reasons of principle supporting OS movement; Comparable switching effort to upgrading to newer version of MS Office; Open file format / standards; More supported languages / localizations ; Smaller document size; Multi / cross-platform support; Easier to customize; OS license desirable; Lack of anti-piracy harassment features; Easier to learn and use; Lower deployment and testing time
Total	**45 product-specific factors identified in total**

Source: MacAulay, M. (2009)[10]

It was further hypothesized that users are not a uniform group and that different types of users might respond differently. In order to identify the factors that might differentiate users, an extensive review of the consumer marketing literature was conducted, and a list of user traits that affect product adoption decisions was developed. The following table (Table 2) describes these user factors.

Table 2: User characteristics from consumer marketing literature

	Consumer behaviour factors theorized to affect product adoption
	Users' ethical considerations related to software piracy
	Users' social class
	Users' innovativeness
	Users' loyalty to Microsoft
	Users' level of cultural identification with the open source movement
	Users' product involvement with Microsoft Office
	Users' dissatisfaction with Microsoft Office
	Users' tendency to search for information about alternatives to Microsoft Office
Total	**8 user-specific factors identified**

Source: MacAulay, M. (2009)[10]

A questionnaire was developed based on these factors. It identified users of Microsoft Office who had considered switching to OpenOffice.org and asked them to identify the product factors that lead to the decision

ISBN 978-0-9566139-1-2

to not switch. It further classified the users based on their characteristics to differentiate their switching behaviour. The following figure (Fig. 1) shows a selection of questions from the questionnaire.

Using the scale, please indicate how much you agree that each statement is a reason why you have not switched from Microsoft Office to OpenOffice.org. If you are not sure, or prefer to not answer, please circle "N/A".

		Completely disagree			Neither agree or disagree			Completely agree	
HI-A:	I am comfortable with Microsoft	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(N/A)
HI-B:	I would not be accepted into the OpenOffice.org culture	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(N/A)
HI-C:	Microsoft offers discounted versions of Microsoft Office	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(N/A)
HI-D:	I was told to not switch from Microsoft Office to OpenOffice.org	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(N/A)
HI-E:	I do not believe in the principles of the open source movement	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(N/A)
HI-F:	OpenOffice.org does not meet my language or localization needs	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(N/A)
HI-G:	I have no interest in building or maintaining an open source community	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(N/A)
HI-I:	Microsoft better understands my needs	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(N/A)

Figure 1: Select questions from questionnaire

RESULTS

259 questionnaires were filled out and returned. 240 were usable after imputation. Respondents were fairly well distributed across standard demographics, including gender, age, household income, and level of education. 84% of respondents indicated that they used Microsoft Office as their primary office suite. Of these 84% of respondents, half of them indicated they had considered switching to using OpenOffice.org as their primary office suite.

Comparison of the responses of the considered switchers and those who had not considered switching office suites revealed that three product-specific factors and three user-specific factors were positively correlated to the choice to not switch office suites. The following table (Table 3) shows the significant factors.

Table 3: Significant product-specific and user-specific switching barriers

Product-specific switching barriers	User-specific switching barriers
Comfort with Microsoft	User innovativeness
Habituation to Microsoft Office interface	Loyalty to Microsoft
Dependence on Microsoft Office file formats	Likelihood to have actively searched for information about alternatives to Microsoft Office

Source: MacAulay, M. (2009)[10]

DISCUSSION

Product Characteristics

Of the 45 product-specific factors identified in the technical and trade press, only 3 factors were recognized by the users as being related to their switching decision. This result shows a disparity between the espoused beliefs

ISBN 978-0-9566139-1-2

in the community and the actual factors that affect a user's decision-making process. There is a significant misunderstanding of the user perspective.

Comfort with Microsoft

The first significant characteristic was comfort with Microsoft. As suggested in the literature,[11] the results show that users who report having considered switching from Microsoft Office to OpenOffice.org did not end up doing so, in part because they are comfortable with Microsoft. The result raises the question of what it is that users are comfortable with. It may be that they are comfortable with the brand due to Microsoft's substantial investment in brand development for their company and products. Brand visibility and consumer recognition improves consumer perception of brand quality and affects their purchasing decisions with conventional products and services.[12] It is likely that the same effect is present in the case of software adoption.

It may also be that users perceive Microsoft to be the 'expert' in the field of Office suites due to its leading market position. Research has shown that expert opinions increase consumer comfort with a product or service.[13] It may be that users feel comfortable with Microsoft because they believe that Microsoft must have the best product offerings available. It is further possible that the concept of 'comfort' takes on several different dimensions. Perhaps users are comfortable with things that they associate with Microsoft, such as the look and feel of their software. This notion is supported by the partial correlation between the first factor and the second significant factor, habituation to the Microsoft Office interface, suggesting that comfort with Microsoft stems in part from familiarity with the interface.

The best strategy to deal with this issue may be to view the competing offering as a supplement, instead of a replacement to Microsoft's products. Users cling to Microsoft and its products out of a fear of losing the comfort they have developed in their day-to-day activities. A radical change is likely to be unwelcome to users. Instead, a gentle introduction of a competing product, with the promotion of its unique features and highlighting of its benefits to the user may be more successful. For example, OpenOffice.org supports one-click PDF export, a feature not yet supported in Microsoft Office. By promoting parallel usage of OpenOffice.org, users can begin to use the product when they require the unique features that Microsoft Office doesn't offer. Over time, as they become more familiar with the new product, their dependence on, and comfort with Microsoft's products will decrease, and a full transition to OpenOffice.org could be made more smoothly.

Habituation to Microsoft Office Interface

The second significant characteristic was habituation to the Microsoft Office interface. This finding supports the technical and trade literature.[14] It suggests that once users are conditioned to using a type of software in a particular way, having them relearn common tasks, even with variances that are as small as different menu layouts, is a significant switching barrier.

Interface design is a topic that is frequently discussed in open source communities, and highlighted as a challenge for many projects, including OpenOffice.org. The program mimics the interface of Microsoft Office quite well, and most of the basic features are accessed in the same way. However, many of the more advanced features and even some of the commonly used features have a different interface. Users that are new to office suites in general find the OpenOffice.org interface to be intuitive as it is designed to be practical and logical; but users that have used Microsoft Office for a long time have become accustomed to its interface. They find OpenOffice.org's interface unintuitive and confusing. This comfort with the user interface appears to overlap with a sense of comfort with Microsoft and its products in general, as there is a moderately strong correlation between this characteristic and the first one.

Many developers have strongly resisted the idea of mimicking Microsoft Office's user interface, arguing that its design is sub-optimal, and that the design decisions have no merit.[15] Yet, it is clear that, from the user perspective, this issue is a significant problem and further modification of OpenOffice.org's user interface to better mimic Microsoft Office may be necessary to attract a larger user market share.

Dependence on Microsoft Office File Formats

The third significant characteristic was dependence on the Microsoft Office file formats. It is well established in the technical and trade literature that users feel dependent on the standard Microsoft Office formats that include

ISBN 978-0-9566139-1-2

'.doc' for word processing documents, '.xls' for spreadsheets, and '.ppt' for slide show presentations.[16] The issue is so pervasive, in fact, that it is listed as one of the primary targets for corrective action in OpenOffice.org's strategic marketing plan.[17]

The sticking point with this switching barrier is that, by and large, OpenOffice.org fully supports the Microsoft Office file formats. The annoyance for users is where there are subtle variations in the way the OpenOffice.org renders Microsoft Office files, leading to small changes in layout. Developers that understand the technical challenges are more forgiving of this limitation, and, in some cases even blame Microsoft for creating poor file format with closed standards in the first place. Users, on the other hand, just want files to look the same across programs, and don't care why or how. A strategy to deal with this issue is to create a 'compatibility mode' in OpenOffice.org that changes its default behaviour to behave exactly like Microsoft Office. In this mode, the file creation, and file handling would behave as 'incorrectly' as Microsoft Office, but produce the results that users expect based on their past use of Microsoft Office. Portability of files between Microsoft Office and OpenOffice.org would be guaranteed to be exact, even if it were considered technically imperfect. While some members of the OpenOffice.org developer community might find this solution distasteful, it may be necessary in order to increase user adoption.

Another approach to addressing this issue is to get users to adopt different file formats. OpenOffice.org has put substantial marketing effort into promoting open file formats such as OpenDocument Text, or '.odt'. But adoption of the formats has been slow. Even with large corporate backers such as IBM, and government support from many European Union countries, the standard still has a long way to go before it becomes ubiquitous.[18]

Another point of annoyance for end users is the default file format setting. The default file format setting in OpenOffice.org is set to OpenDocument formats. Users must manually select a different file format setting if they want to save a document in a Microsoft Office file format. Many users cannot be bothered to change the default setting, or get confused by the meanings of file extensions and file types, and simply assume that there is limited or no support. Some research has suggested that changing the default file type setting to the Microsoft Office formats preemptively, such as at install time, can reduce user frustration with the product.[19]

Additional Insights on Product Characteristics

The results on product characteristics related to adoption were somewhat unexpected. Another unexpected result was that half of the users of Microsoft Office who answered the questionnaire reported that they had considered switching to using OpenOffice.org as their primary office suite. This number should be very encouraging for open source proponents as it shows that competitors are making inroads into Microsoft's mindshare when it comes to office products, where it previously enjoyed a near-monopoly. There has not previously been an investigation of what users were thinking with regards to switching considerations. All previous research has looked purely at adoption numbers leading to suggestions that there is a big difference between the products and that the difference leads to lower user adoption of OpenOffice.org. The present findings suggest that the differentiation between office suites, from the user perspective, may not be as large as previously thought.

When considered from an interdisciplinary perspective, using the lenses of consumer marketing, this finding is perhaps not surprising. Users care about different things than developers. Their goals in using the software are different, and, as such, their concepts of value are different than a technology-focused developer. It is quite possible that, from the user valuation perspective, there is very little differentiation between office suites. Users think of Microsoft Office vs. OpenOffice.org in the same way they consider Coke vs. Pepsi. They don't care about the specific ingredients; they care that the drink relieves of their thirst. McGrath and MacMillan describe the analogy in the car manufacturing industry, explaining the sort of challenge that this presents for open source development communities:

> *Given little else to differentiate cars in a class, a person may think, why not buy the one with the cup holder? A cup holder makes it a lot easier to drink your morning coffee while driving to work. This kind of consumer purchasing behaviour can be the despair of engineers, designers, and scientific staff. They spend their lives worrying about things like fuel efficiency, advanced hydraulics, and wind resistance.*[20]

ISBN 978-0-9566139-1-2

User Characteristics

The research on product characteristics made it apparent that users have a very different perspective from developers when it comes to what they value in office suites. Through the user's eyes, the differentiation between products is not as broad as previously thought. This finding puts focus on the characteristics of users that affect switching decisions, independent of product-specific characteristics. These characteristics are of particular research interest as they may reveal strategies for improving adoption by directly appealing to the user, instead of changing the product to suit them.

User Innovativeness

The first significant characteristics was user innovativeness, which was defined as the tendency to willingly embrace change, try new things, and adopt new products more often and more quickly than others.[21] This finding is as expected. It supports the notion that more innovative users view the risks of switching to be much lower, and hence are more eager to adopt novel products.

Looking through the lenses of consumer marketing theory, two strategies are likely to have an effect on user switching behaviour. The first is to appropriately target marketing according to user innovativeness. More innovative users are likely to respond better to appeals to the technical benefits of switching to OpenOffice.org. They are also likely to view the risks of switching to be lower. Traditional product-specific marketing that promotes the value metrics that these users are interested in will likely have the expected effect. On the other hand, less innovative users will not respond well to technical benefit-oriented marketing. Instead, marketing that focuses on simplicity, lower cost, and ease of use may appeal more to less innovative users.

The second strategy is to remove the switching consideration altogether for target markets with users with low innovativeness. One effective way would be to focus on pre-installation on new computers, or on bundling with other applications. The key to the success of this strategy is to engage first-time users of office suites, instead of users that switch from Microsoft Office. But, given the length of time Microsoft Office has been available, it is challenging to find users that have not used it. A possible approach would be to focus on adoption in recently industrialized nations. A significant portion of the world's population has not yet entered the computer age. As computer usage continues to increase around the world, targeting countries with a rapidly expanding infrastructure, such as China or India, may be a more effective strategy to increase adoption than taking Microsoft head-on in their established footholds in North America. This approach is one of the strategies that the OpenOffice.org Strategic Marketing Plan has prescribed and is currently implementing.[17]

Loyalty to Microsoft

The second significant characteristic was loyalty to Microsoft, which was defined as the strength of the relationship between an individual's relative attitude towards an entity and repeat patronage.[22] This finding is as expected, supporting the notion that users who had considered switching office suites had lower loyalty to Microsoft.

Microsoft has done an excellent job at developing and maintaining user loyalty towards the company and its products. Loyalty can be very difficult to overcome. One strategy to address the issue is to identify customers who may be potential defectors and focus on them. Happy, loyal customers are extremely difficult to draw away – so much so, in fact, that resources are probably best not spent on them. It may be far more effective to target the once-loyal users who, for one reason or another, have begun to consider defection.

A strategy that may prove effective is to search for the root causes of customer departures in order to identify the failure points in the competition. By focusing marketing on these points, and providing an alternative that resolves the issues with the incumbent's product, new products may be able to pick up market share. A good approach may be to target users that are overserved by Microsoft Office, i.e. users who feel that it is too expensive, that they don't use all of its features, and that they are not willing to pay more for additional features or improvements on existing features. Such users are good candidates for a targeting with a disruptive product that offers comparable basic features at a lower price, as is the case with OpenOffice.org.[23]

ISBN 978-0-9566139-1-2

Likelihood to have actively searched for information about alternatives to Microsoft Office

The third significant characteristic was likelihood to have actively searched for information about alternatives to Microsoft Office. This finding is expected, supporting the notion that users who actively search for information about alternatives are more likely to have considered switching.

This characteristic should be considered more lightly than the previous ones as only its correlation with switching consideration has been established. There is no evidence to support a causal relationship. It is possible that users who are already considering switching search for information on how to go about making such a switch. One should be cautious in assuming that the opposite may be true.

Nevertheless, were there a causal relationship, the course of action would be to promote the availability of information about alternatives to Microsoft Office, a practice that is well under way. The key in this strategy is getting information to the masses of users who wouldn't otherwise have come across it. Getting a mainstream user base interested can be challenging. Sun Microsystems has taken some steps in this direction by including advertisements for OpenOffice.org in products that mainstream users are more likely to come across, such as the Java runtime that is installed on most Windows computers. When the Java software runs its scheduled update, as the update downloads, the user is presented with a small ad that lists the value proposition of Microsoft Office, including that it is available at no cost, compatible with Microsoft Office, supported on numerous operating systems, and has a one-click PDF export feature. The brevity of the message encourages users to do additional research on the product to learn more. It is this sort of minimalist yet widely distributed message that will likely help expose potential users to the research material that could promote switching behaviour.

Notable factors that did not achieve statistical significance

The findings on user characteristics that affected switching consideration were expected. They support similar findings in the consumer marketing literature on product adoption. Also notable is which user characteristics did not yield a significant result, especially where the literature suggested that they should. Of the eight user characteristics that were assessed, five were not found to be significant predictors of switching consideration.

One such example is the measure on dissatisfaction with Microsoft Office. It was not found to be significantly different between users who reported that they had seriously considered switching, and those who did not. This result is unexpected. It appears to contradict the extant literature on the effects of dissatisfaction on product switching behaviour. The classical assumption is that dissatisfaction with a product leads users to seek out alternatives, with which they might be more satisfied.[24] A possible explanation for this result is that, in the case of office suites, the initial motivation to consider switching is not related to dissatisfaction, but rather is related to another factor. Anecdotal evidence supports this possibility as several respondents indicated that it didn't matter whether or not they liked their office suite as they had no choice in the office suite they use. It may be a necessary factor that users feel that they have the ability to switch products of their own accord in order for dissatisfaction to be a switching motivation.

Another absent effect was that of user cultural identification as a switching motivator. It is widely believed in open source communities that identification with the ideals of the community is a reason people use the software.[25] The results did not support this hypothesis. A possible explanation is that the concept of culture in open source community is markedly different from traditional measures of culture such as ethnic or geographic cultural identification. The open source community is made up of widely varying groups, from all around the world, with disparate objectives, and different forms and venues of representation. It is possible that cultural association forms differently in the case of the open source community, and that users experience it differently from other members of the community.

Additional research that looks specifically at the factors that did not achieve significant results may be warranted to shed light on these apparent anomalies and help generate novel models and measures.

Additional Insights for Open Source Communities

By using an interdisciplinary approach, the contributions of users to open source ecosystems become a lot more explicit. There are numerous lessons to be learned from user adoption behaviours and motivations and these lessons cannot be clearly understood from a purely technical perspective. In order for open source development

ISBN 978-0-9566139-1-2

communities to make inroads in user adoption the gap between the views of developers and the views of users must be reduced. The best way to accomplish this goal is likely through increased user engagement. While a primary motivation for developers to participate in open source software development is to scratch a personal itch[1] – and this is a laudable goal – developers must consider the potential benefits of an increased user base on the long-term viability and relevance of the product to which they are contributing their time. Characterizations of users as leeches who do not contribute to open source projects push away an important resource.

Users are a key part of open source ecosystems. They provide tremendous amounts of free advertising through word of mouth and other viral effects, advertising that brings in new developers along with new users, and, occasionally, even corporate sponsorship. The health of an open source ecosystem is best quantified by the strength of its network,[26] and users *are* that network. Further, users are an important Geiger counter for failure. The number and behaviour of users can be important metrics to understand mistakes, poor product design decisions, and incorrectly executed strategies; and, most importantly, they provide the means to measure the effectiveness of corrective action.

There is no shortage of user-based communities. The traditional model has seen open source ecosystems grow around developers. Users were expected to contribute to developer-focused resources such as mailing lists, and bug tracking databases. Users feel intimidated in these environments, and so do not contribute. A more successful approach may be to turn around this model, and get developers to visit and participate in user communities and engage users on their turf, where they feel comfortable. Not only will this approach put users at ease, and help draw out better feedback from them, but also the environment itself is a telltale sign of how users aggregate around a piece of software. It provides hints at ways users want to interact with the product, and what they value about it, and could provide ideas for future design.

CONCLUSION

This paper examined the switching behaviour of users in the case of office suites. The results showed that there was a significant difference between espoused beliefs of community practitioners and the switching motivations of users. The lessons learned are that open source communities must adjust their approaches in order to pay more attention to the user perspective and user value metrics if they wish to improve adoption of their products. The marginalized user contributes to open source communities as surely as other participants, but in different ways. These contributions are essential to the health of the ecosystem and present an opportunity for improvement and adoption growth.

Traditional consumer marketing metrics can be used to learn about users' behaviours and contributions. By better understanding the users, open source communities can create more useful products that better meet user needs. Different types of users behave differently, and it may be possible to encourage them to participate in ways that relate to their interests. For example, innovative users tend to adopt technology more readily, don't mind bugs and crashes as much, and are willing to put in the time to help report errors and suggest improvements. Loyal users may not be technically savvy, but will gladly wave the banner of the company, promoting the product far and wide and bringing in new users. There are many other passive participant roles that users can play when not marginalized, such as event promoters, designers of complementary products, documentation creators, and financial donors, all of which help improve the health of the open source community.

REFERENCES

[1] Bonaccorsi, A., & Rossi, C. *Comparing motivations of individual programmers and firms to take part in the open source movement: from community to business.* (2006) Knowledge, Technology and Policy. 18(4): 40-64.
[2] Henkel, J. *Open source software from commercial firms: tools, complements, and collective intervention.* (2003) Working paper. Munich, Germany: Institute for Innovation Research, Technology Management, and Entrepreneurship, University of Munich.
[3] Lerner, J., & Tirole, J. *The open source movement: key research questions.* (2001) European Economic Review. 45: 819-826.
[4] Snyder, B. *The fight over open source 'leeches'.* (2009) InfoWorld. June 1.

ISBN 978-0-9566139-1-2

[5] Scipion, F. *Understand Open Source ecosystems.* (2007) Madrid: 55thinking.
[6] Proffitt, B. *The Myth of the Freeloading User.* (2009) Linux.com. June 1.
[7] Koenig, J. *Seven open source business strategies for competitive advantage.* (2004) IT Manager's Journal.
[8] Pirillo, C. *Users vs. Developers.* (2006) Accessed at: http://chris.pirillo.com/users-vs-developers/
[9] Colgate, M., & Lang, B. *Switching barriers in consumer markets: an investigation of the financial services industry.* (2001) Journal of Consumer Marketing. 18(4): 332-347.
[10] MacAulay, M. *Assessing the switching barriers between Microsoft Office and OpenOffice.org.* (2009) Master's Thesis (Unpublished). Ottawa, Canada: Carleton University.
[11] Scoble, A. *Is Microsoft Office too expensive? Not if you need it.* (2006) Computerworld Blogs.
[12] Wulf, K. D., Odekerken-Schröder, G., Goedertier, F., & Van Ossel, G. *Consumer perceptions of store brands versus national brands.* (2005) Journal of Consumer Marketing. 22(4): 223-232.
[13] Aqueveque, C. *Extrinsic cues and perceived risk: the influence of consumption situation.* (2006) Journal of Consumer Marketing. 23(5): 237-247.
[14] Matzan, J. *Making the leap: Microsoft Word to OpenOffice.org Writer.* (2005) Linux.com. April 20.
[15] Haff, G. *When choice is bad: The OpenOffice ribbon.* (2009) The Pervasive Data Center. CNET News. August 10.
[16] Dolinar, L. *What's holding OpenOffice back?* (2008) LinuxInsider. July 7.
[17] OpenOffice.org. *Strategic marketing plan.* (2008)
[18] Martens, C. *IBM undeterred by setbacks to ODF adoption.* (2007) InfoWorld. June 8.
[19] Rossi, B., Scotto, M., Sillitti, A., & Succi, G. *Criteria for the non invasive transition to OpenOffice.* (2005) Proceedings of the First International Conference on Open Source Systems. Genova. 250-253.
[20] McGrath, R.G., & MacMillan, I.C. *The entrepreneurial mindset: Strategies for continuously creating opportunity in an age of uncertainty.* (2000) Boston, MA: Harvard Business School Press, p. 28.
[21] Hirunyawipada, T., & Paswan, A. *Consumer innovativeness and perceived risk: implications for high technology product adoption.* (2006) Journal of Consumer Marketing. 23(4): 182-198.
[22] Donio, J., Massari, P., & Passiante, G. *Consumer satisfaction and loyalty in a digital environment: an empirical test.* (2006) Journal of Consumer Marketing. 23(7): 445-457.
[23] Christensen, C. M. *The innovator's dilemma.* (1997) Boston, MA: Harvard Business School Press.
[24] Jones, T. O., & Sasser Jr., W. E. *Why Satisfied Customers Defect.* (1995) Harvard Business Review. November-December.
[25] Miller, R. *How OpenOffice 2.0 stacks up against Microsoft Office.* (2006) SearchEnterpriseLinux.com.
[26] Moore, J.F. *The Death of Competition.* (1996) New York, NY: Harper Business.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

ELASTIC HOLOGRAMS

Juan Leonardo MARTINEZ-HURTADO

PhD Student, University of Cambridge, Cambridge, United Kingdom
juanleonardo@gmail.com

Keywords: Hologram, Elastic, Silicon, Grating

ABSTRACT

A new technique of making holograms is used to produce gratings in elastomeric materials. A holographic grating contains silver nanoparticles arranged in ordered fringes or layers separated by half of the reflected wavelength. These nanostructures act as photonic filters that reflect only certain colors at a defined angle. The materials used to construct the holograms are polymeric elastomers that transduce a change in internal conformation to a variation of reflected color, thus changing hologram's coloration when deformed, stretched or compressed. We have shown how a change in applied force or pressure can be directly related to a change in color. These devices can be applied in construction materials, machinery, coatings, etc.

INTRODUCTION

Holography is a technique in which a coherent light (i.e. one wavelength)[1] is used to generate a grating embedded in a material. Such grating is recorded in depth inside thin polymer films. The final outcome or product is, as in black and white photography, a flat surface containing the recorded information carried by the reflected light. Using photography as an example, the same principles that generate a change in the photographic films are used in holographic films. Silver salts are reduced to form silver nanoparticles that generate patterns or fringes inside the polymer matrix containing them. In the case of holography, a coherent light source is used, meaning that only one wavelength is present; therefore, it is possible to form standing waves that form nodes and antinodes with different energy profiles. The energy in the antinodes is enough to reduce photosensitive silver salts. In this paper, another method is discussed where the fringes are formed in the anti-nodes instead. A simple approach of this explanation is depicted in Figure 1.

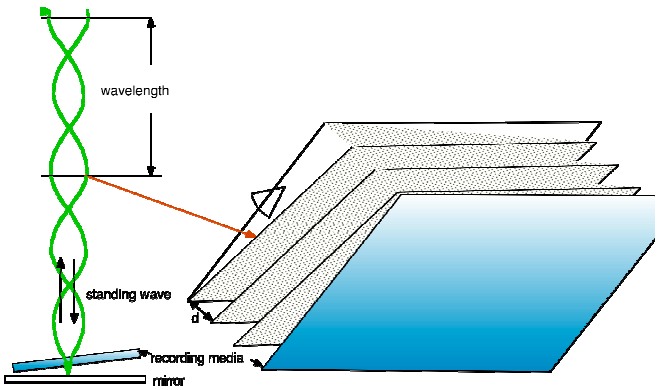

Figure 1: Formation of holographic volume gratings in a confined space. The drawings are not a scale. Laser wavelengths are usually about hundreds of nanometers.

ISBN 978-0-9566139-1-2

Using certain arrangements it is possible to record reflections from objects and it is possible to store three-dimensional information of the object in confined spaces.[2]

When recording a hologram, the inclination of the recording media towards the reflecting surface of the mirror gives a different angle to the normal angle of incidence, and that way an external white light source can be used to see the holographic colors. This is due to an interesting photonic effect generated by such pattern or periodicity at the nanoscale (half or fourth of the wavelength).[3] The wavelengths are filtered, as they are reflected by the fringes formed in the hologram, this construction is also called Bragg reflectors, and can be theoretically analyzed as simple thin film filters[4] or 1D photonic crystals.[5] Another simple way of analyzing them is proposed by Kogelnik,[6] an analogy of the treatment proposed by Bragg for diffraction of X-rays in crystal lattices, hence the equation is commonly mistaken as the Bragg's law because it contains similar terms.

$$\lambda = 2nd\cos\theta$$

Where λ is the peak wavelength, n the average refractive index of the material, d the space between the fringes and θ the angle at which the Bragg condition is satisfied. A graphical explanation of this is depicted in Figure 2 with the original expression.

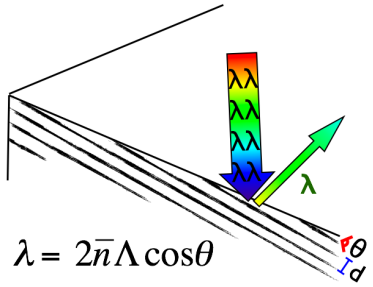

Figure 2: Photonic effect observed when light hits the holographic grating: only certain wavelengths are reflected and the rest are filtered; the equation in the figure corresponds to the original expression for studying such gratings, the grating spacing *Λ* is also expressed as *d*.

Elastic holograms

Holograms are usually made in hydrophilic polymer films, traditionally gelatin films, in which silver salts in aqueous solution homogenously mix with the film materials, thus allowing the gelatin to contain enough salts to be reduced when exposed to laser light. Other polymeric materials can also be used; the majority are hydrophilic films so that aqueous solutions can permeate. Recently a method was developed to incorporate gratings in hydrophobic materials,[7] such as silicon, opening up the possibilities of applications for volume holographic gratings. In this case the particles are formed first homogenously distributed in the film and then using the same principle some are cleared off with coherent laser light energy; this of course requires more energy to form the fringes; contrary to normal holography the fringes are formed in the nodes. Usually the holograms built in that way are designed for fabrication of sensors but also mechanical changes can be monitored when using the right material.

METHODOLOGY

Hologram construction

The holograms were constructed in a polymeric film made of silicon (poly-dimethylsiloxane, PDMS). The PDMS films are prepared as stated by the fabricant, mixing 10mL of pre-polymer solution Sylgrad 184 ® (DowCorning) and 1mL of curing agent containing a Pt based catalyst. The mixture is then deposited on top of a flat transparent surface and was left to cure overnight. The silver particles were incorporated as described in the recent work Martínez-Hurtado et al[8] and then exposed to laser radiation to form the grating, resulting in a shiny green elastic hologram.

ISBN 978-0-9566139-1-2

Hologram interrogation

The hologram, was interrogated by stretching and compressing it. The applied force used for stretching was calculated using $F=ma$; where F is the force, m is mass, a corresponds to acceleration, in this case only gravity was used, therefore 9.81 m/s^2. F was modified by adding heavy objects of known mass to the device in Figure 3.

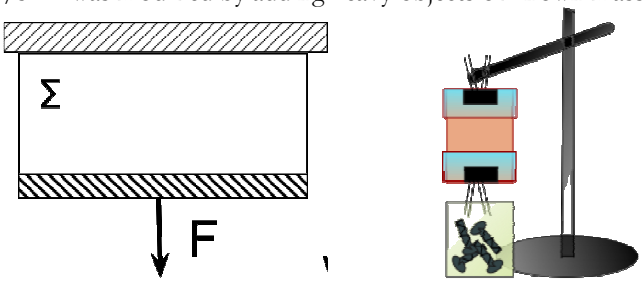

Figure 3: Force diagram of the interrogated hologram and stretching device.

When stretching or compressing the hologram a change in conformation of the internal structure also occurs, therefore, the distance between the fringes is expected to change and a change in reflected color is also expected. The materials in the hologram are not changing; the refractive index is therefore not changing and the angle of observation can be maintained fixed so that the only parameter changing is the distance or spacing d. A slice of the hologram in both stages would look as that in Figure 4.

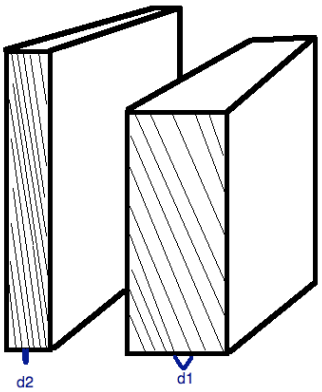

Figure 4: A short caption describing the figure.

With the help of optic fibers connected to a spectrophotometer that detects photon counts at different wavelengths, it is possible to maintain a constant angle from the incident light coming out from one of the fibers and the fiber for the detector the fibers where placed in front or below according to the set up, so that it is possible to detect the reflected wavelengths generated at different times. In a similar fashion the hologram was compressed as in Figure 5, and the wavelength measured by changing the optic fibers position placed underneath the hologram at the right angle. The pressure was calculated as $P=F/A$ where P is pressure, F is force and A is the area. The force was calculated as in the previous example by adding heavy objects over a 1cm diameter circle, and then the pressure was calculated.

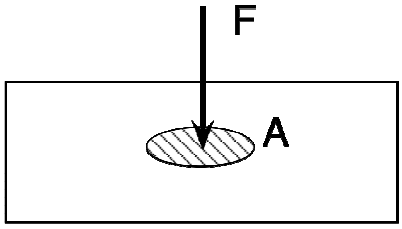

Figure 5: Force diagram of the compressing device used.

ISBN 978-0-9566139-1-2

RESULTS AND DISCUSSION

The intensity of the holographic reflection as well as the peak wavelength were recorded at different times when a force was applied. The change is reported as the wavelength change $\Delta\lambda$ the result is plotted against applied force in Figure 6. It is noticeable that the intensity decreases as the wavelength decreases. When stretching the hologram the space between the particles increases, therefore the separation allows more light to go through and less light to be reflected resulting in a decrease in intensity as shown in the figure.

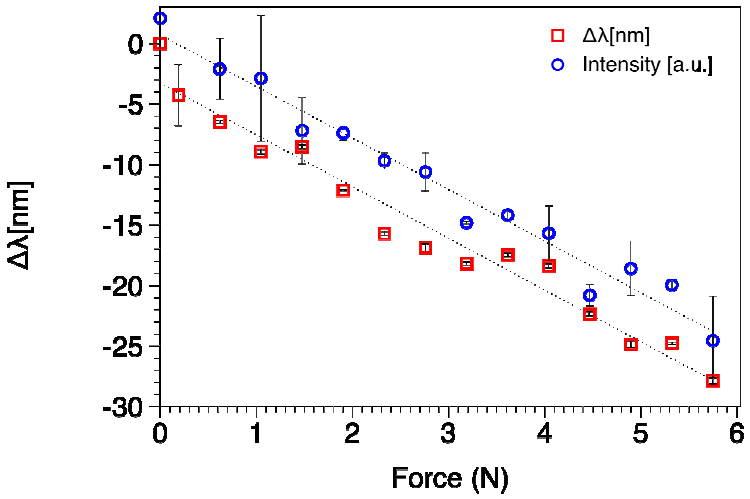

Figure 6: Stretching the hologram results in a contraction of the fringe spacing therefore a decrease in wavelength and intensity.

A similar result is obtained when compressing the hologram: the intensity and wavelength peak position decreased. The graphs in Figure 7 show this phenomenon.

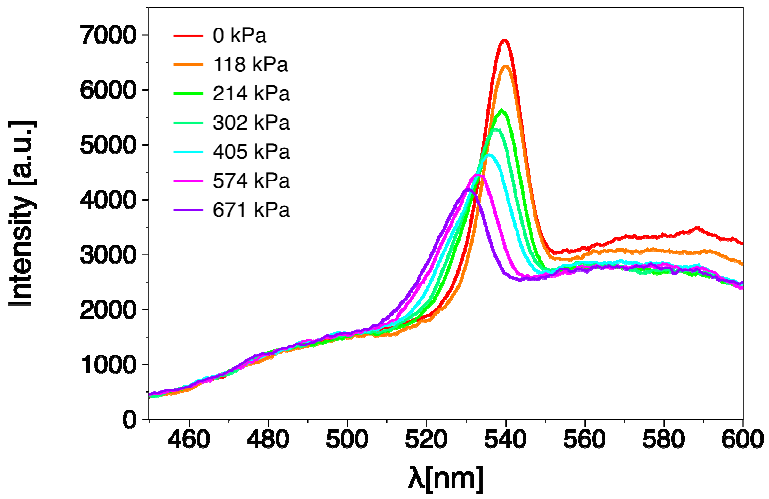

Figure 7: Visible spectrum of a hologram being compressed with certain pressures.

When repeating the experiment through time, it was noticed that the reflection and wavelength peak were not in the same position as when the heavy objects were loaded. This is due to a hysteresis effect. The polymer films

ISBN 978-0-9566139-1-2

take longer to relax and decompress to its original position. Figure 8 shows that effect when different pressures where applied to the hologram.

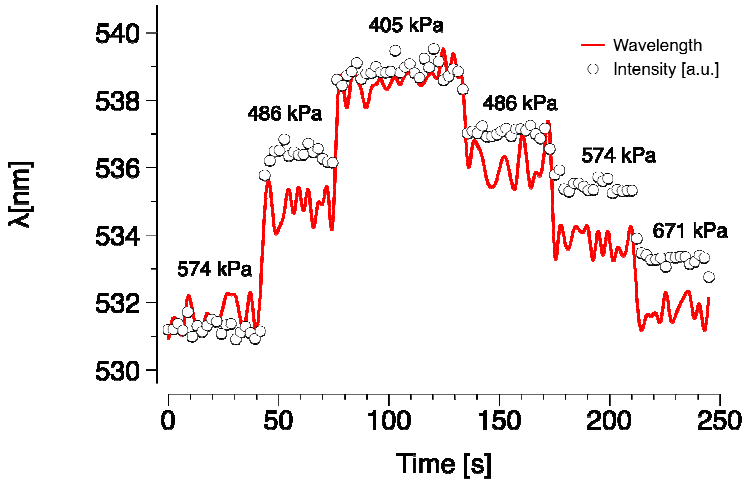

Figure 8: Compressing hologram hysteresis effect when compressed and decompressed sequentially, loads added in kPa.

Having a pressure or tension sensitive hologram can provide an opportunity to use it as a pressure sensor or stress sensor in materials or coatings covering certain materials. Recording images that change when stretched or pressurized can also be achieved with the right construction, thus a display can be designed using this principles.

CONCLUSION

An elastic hologram that changes color when stretched or compressed was fabricated and analyzed. The intensity and wavelength peak can be related to the pressure or force applied. When the fringe spacing is changed mechanically the reflected color and intensity change accordingly.

REFERENCES

[1] A laser is a source of coherent light that only emits at one wavelength. The one used in this study is a 532nm pulsed laser.
[2] In other words it is having a 3D reconstruction flattened to the nanoscale where the 3D features are constrained to the thickness of the polymer film, which usually is in the range of 10-2000μm.
[3] Vukusic, P. *Natural Photonics.* (2004) University of Exeter. IOP Physics World.
[4] Mc Leod, J. (2006) *Thin film optical filters*, 3rd Ed. Taylor and Francis.
[5] Joannopoulos, J.; Johnson, S.; Winn, J.; Meade, R. (2008) *Photonic crystals: molding the flow of light*; Princeton University Press.
[6] Kogelnik, H. (1969) *Coupled wave theory for thick hologram gratings*. The Bell System Technical Journal. 1969, 48, p2909.
[7] Martinez-Hurtado, J. L.; Davidson C. A. B.; Blyth J.; Lowe C. R.; (2010) *Holographic sensors for hydrocarbon gases and other volatile organic compounds*, (submitted for publication).
[8] Idem.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

THE SUBTEXT OF NEW HUMAN RIGHT CLAIMS: A Socio-Legal Journey Into the 'Right to Truth'

Patricia NAFTALI

PhD Student (FNRS Research Fellow), Université Libre de Bruxelles (ULB), Belgium
pnaftali@ulb.ac.be

Keywords: Enforced Disappearances, Sociology of Human Rights, Transitional Justice, Grassroots Activism, Social Movements, Truth Commissions, Amnesties

ABSTRACT

While historians, philosophers and legal scholars have long debated the distinction between historical and judicial truth, it may come as a surprise that a 'right to truth' has recently emerged in international law in the context of serious atrocities. After three decades of struggle, victims' advocates have recently gained momentum with the 2006 United Nations' unprecedented adoption of a treaty on enforced disappearances that expressly recognizes their 'right to know the truth' about what happened. This legal innovation raises the question how victims' groups came to successfully voice their claims to the world community. However, positivist legal scholarship traditionally overlooks the social dynamics by which norms emerge, thereby failing to account for how moral and political claims migrate from the margins to become express legal norms. Through the case study of the right to truth, this paper traces an itinerary from its local formulation and reception, to its international recognition, to illustrate how a socio-legal perspective might enrich our understanding of the creation of human rights norms, by uncovering the role of grassroots movements and transnational advocacy strategies, discursive mobilizations and professional trajectories or biographies. It explores the role of non-state actors in formulating novel human rights claims, the importance of moral and cultural resonance in issue-framing, and the ongoing dialogues between the local, regional and international spheres.

"The people have a right to the truth as they have a right to life, liberty and the pursuit of happiness." – Epictetus

"[T]he expression 'to have a right to truth' is meaningless. One must say, rather, that man has the right to his own truthfulness (veracitas), i.e., to subjective truth in his own person." – Immanuel Kant

INTRODUCTION: Mapping an Itinerary of A legal innovation

Could 'truth' possibly be subject to a right of its own? Contrary to a traditional natural-law concept of truth as a necessary element of justice, the quest of establishing the 'truth' might be considered an illusion, and possibly a dangerous one. Notwithstanding the ambivalence such a concept entails, the notion of a 'right to truth' owed to victims of gross state abuse has in recent years become salient in political and legal discourse, with varying subtexts ranging from the opening of investigations into state-sponsored abuse, victims' access to justice, declassification of secret archives or reparations. At the international level, if the importance of truth in peace-resolution or post-conflict rebuilding features high on the agenda of the United Nations (*UN*), it seems to have recently shifted from a moral imperative to a legal one, as illustrated by the recent adoption of the *International Convention on Enforced Disappearances* that expressly recognizes to victims a right to truth.[1] In this context, how did relatives' *'desire to know the fate of loved ones'* characterized by the UN General Assembly in 1974 as a '*basic human need which should be satisfied to the greatest extent possible,*'[2] become an express '*right to know what happened to their next-of-kin*'[3] three decades later, in 2006?

ISBN 978-0-9566139-1-2

This recent consecration and the increasing institutionalization of the right to truth in international and regional bodies raises the question of why and how do certain claims formulated at a national level succeed in permeating the international legal realm, while others fail.

In this regard, social constructionist scholars, who view rights as '*an artefact produced through social processes of framing and construction*',[4] have advanced some interesting insights to explain how social issues grounded in historically-specific contexts are translated into human rights disputes at the international level. Although human rights sociology is still recent, Ken Plummer suggests for instance that *'the task for sociologists is to become intimately familiar with the crusaders, their claims and the social processes though which rights emerge'*.[5] In particular, authors such as Martha Finnemore & Kathryn Sikkink have tried to elaborate a model that accounts for the complete life span of an international norm, from its formulation as a domestic norm to its endorsement by formal actors of the international system.[6] They identify three stages of a '*norm life cycle*': norm emergence, during which various social actors deliberately construct and promote a norm understood as a standard of behaviour that they deem appropriate; norm institutionalisation, that occurs if a critical mass of followers accept and start promoting the new norm in turn; and finally, norm internalisation, where the norm is taken for granted. They further suggest that each stage responds to different behavioral logics. During the first stage of the norm life, '*norm entrepreneurs*' try to gather critical support for change by processes of persuasion and framing; if they succeed, a '*tipping point*' occurs, where broad acceptance of the norm by states or international organizations leads to a '*norm cascade*' through an accelerated dynamic of imitation among new norm followers.[7]

Such a socio-legal framework implies shifting away from a purely legal positivist perspective in two ways. First, it departs from a state-centric paradigm of norm-production that conveys a static view of norm creation. Instead, it invites to look at human rights creation through an actor-oriented perspective to understand who engages in norm creation and what are their motivations. Second, instead of focusing on formal sources of legal creation, such as texts emerging from public institutions such as legislatures and courts, it dwells on informal discursive mobilizations and other strategies of persuasion used by these entrepreneurs to convince other agents to embrace the new norm.

Turning to the right to truth, who have been its main norm entrepreneurs and what has made their claims compelling both internally and internationally? What mechanisms have driven a nascent idea rooted in local demands into a state-supported and commonly accepted institution, to the extent that truth-production has today become a featured instrument of conflict-resolution and rebuilding practices in conflict-ridden states around the world?

To answer these questions, I will draw on some of social constructivists' insights into international norms' dynamic to suggest one itinerary of the right to truth from its formulation by marginalized groups at the grassroots-level onto the international stage. This itinerary attempts to account for the international reception of the right to truth by focusing on the role of non-state actors in legal innovations, their motivations and their strategies, including the various subtexts that accompany their official discourse.

Going back to the norm entrepreneurs that have originally claimed truth against the background of state-sponsored clandestine repression throughout Latin America from 1970s to 1990s, it first highlights the local contribution of grassroots movements in claiming and framing a right to truth in the context of enforced disappearances. To this end, it will focus on a few discursive strategies of family-based grassroots movements to show how they invented a new mode of protest against clandestine state repressive practices that has resonated in both the national and international scenes.

It then examines how the right to truth has migrated from a local context onto the international stage through a process of institutionalisation in various multilateral organizations. More specifically, it shows how various truth entrepreneurs have reframed the right to truth to the point that international and regional multilateral organizations have gradually integrated the right to truth in their broader agendas for peace-making and the fight against impunity. To this end, it focuses on three underlying factors of consolidation or legitimation of the right to truth in the international arena: the development of transnational advocacy networks, the scientific validation of truth-seeking initiatives through the constitution of a new international expertise, and professional migrations between the advocate, academic and political spheres.

ISBN 978-0-9566139-1-2

FROM GRASSROOTS ADVOCACY TO NATIONAL RECEPTIONS OF TRUTH

This part illustrates the contribution of family-based grassroots movements in claiming 'truth' in their struggle to obtain information on the fate of their disappeared relatives under Latin America authoritarian regimes from the 1970s to the 1990s. It first briefly outlines why truth as a concept has been mobilized by families as a specific frame of resistance to protest against enforced disappearances, and more generally, state clandestine repression in Latin America. It then looks at the discursive strategies mobilized by these movements to frame the demand for truth as a natural right and suggests that their success is mainly due to the promotion of motherhood and respect due to the dead, before turning briefly to domestic responses that have accommodated the demand for truth in the political transitions that followed repressive periods.

A. Families Engage in Resistance against Secrecy (1970s–1990s)

The clamor for truth at the origin of the right to truth has mainly arisen in reaction to the extent of the state-sponsored clandestine repression of opponents throughout Latin America in the 1970s and 1980s. Although enforced disappearances were not a new phenomenon, they were carried out with unprecedented intensity in Latin America from the 1960s through the 1990s. Under the national security doctrine, deeply rooted in civil and military dictatorships in Chile, Paraguay, Brazil, Uruguay, Argentina, and Bolivia, political opponents were radically redefined as '*subversive elements*' of society that '*hid*' within political parties, trade unions, universities and minorities' groups. According to Argentine dictator Videla, '*terrorists are not only those who bear with them a bomb or a gun, but also all those who spread ideas which are contrary to the Western or Christian civilization*'.[8] Their secret eradication was thus justified because of the '*clandestine nature of the enemies*'. The most pragmatic way to leave no trace was to make victims disappear: they would 'vanish' one day, and never reappear. Most often, they would be abducted by disguised agents driving vehicles without licenses, detained in unacknowledged prisons, tortured and secretly assassinated before being buried in clandestine cemeteries or mass graves. Persistent state denial about the disappearance would then complete the confinement of victims to anonymity by eventually denying their very existence. Young people were an important target of the repression. As one mother of a disappeared child reported,

> *this secret destiny has been decided according to norms themselves even more secret, which, by their terrorizing and aberrant nature, involve the complicity of armed forces. Our sons and daughters have been victims of this secret code and we, above everything, receive as our punishment the terrible ignorance of that code.*[9]

In reaction to the systematic silence of authorities regarding their requests for information, many relatives of the disappeared thus started to engage in activism because of their personal experiences of loss. The demand for truth regarding the fate and whereabouts of missing relatives became their rallying cry throughout Latin America to protest against disappearances and systematic destruction or falsification of evidence. Organizations of women such as the Argentine *Mothers* (*Madres*) and *Grandmothers* (*Abuelas*) *of Plaza de Mayo*, El Salvador's *Committee of Mothers* (CoMadres), all founded in 1977, and *Guatemala's National Coordination of Widows* (CONAVIGUA) initially emerged as informal groups of women who met while anxiously searching for news about their missing relatives in Courts, police stations, army barracks and government offices or official prisons. Most of them did not have prior political experience.[10] This engagement of families has been seen as emblematic of the passage of families from the private to the public sphere. In a year, the *Madres* movement dwelled from a dozen of mothers to include hundreds of activists by 1978. By 1981, their weekly demonstrations before the Presidential Palace rallied thousands to hundreds of thousands of protesters.

B. Framing the Demand for Truth as a Natural Right: Life, Family and the Dead

At the stage of norm emergence, constructivist and social movements literature particularly emphasize the role of norm promoters in 'framing' new claims by creating or calling attention to issues '*using language that names, interprets and dramatizes them*'.[11] However, as they suggest, new claims don't appear in a normative emptiness, but rather emerge in a space filled with highly competing norms and perceptions of interest. To become influential and rally a critical mass of supporters, a legal innovation thus needs to *'[resonate] with basic ideas of human dignity shared by most cultures around the world*' so as to transcend singular experiences and appear universal.[12] In this regard, the construction of a right to truth by families appears a striking example of sociologist Waters' suggestion that because '*human rights is an institution that is specific to cultural and historical context just like any other… its very universality is itself a human construction*'.[13] Originally formed to locate their own relatives, how did these women's groups gradually turn into social movements by transforming their individual claims into universal demands '*in the name*

ISBN 978-0-9566139-1-2

of every child of every mother' as an effective counter-narrative to the dominant discourse of national security doctrine?

First, most of these family-based organizations articulated their struggle for truth with a moral imperative by officially rejecting any political affiliation. For instance, one of the *Madres'* mottos was '*We don't defend ideologies; we defend life*.' Their first official announcement in 1977 was: '*We do not ask for anything but the truth*,' in echo to President Videla's statement that '*Nobody who tells the truth (would) suffer reprisals*.'[14]

Second, these grassroots women's groups made their appeal from within the existing cultural values by promoting their family-based identity and by framing disappearances as an attack on family. They challenged the moral and religious legitimacy of the *juntas'* self-described Western and Christian values by using their language and praise of traditional family as a cornerstone of the nation.[15] To this end, these movements initially rejected a feminist agenda to rather insist on motherhood.[16] They thus publicly identified as *mothers*, insisting on their '*raison d'être [being] the struggle for [their] children*',[17] as *spouses* or *grandmothers*, not *women*. They wore white headscarves to recall their children's cotton diapers. They consistently insisted on state obligations owed to families by arguing that the authorities infringed on their right to private life by depriving them of truth and of their relatives.

Third, they further appealed to the moral, cultural and religious register by articulating the need to know the truth about whether their missing relatives were alive or dead with the right to mourn and to bury the dead. For instance, the *Madres* made pilgrimages to Marian devotional sites to evoke the pain of mothers seeking children crucified by a state that pretended to be Christian but prevented families from burying their dead.[18] They further insisted on truth as a natural right by claiming that if their children did not reappear alive, '*their right to a plot of earth to rest in peace, by the simple fact of being born, as well as the right to celebrate the cult of dead relatives, would be profaned*'.[19]

Finally, their campaigns to express their anguish publicly captivated world opinion. They emphasized the underlying inhumanity of state silence by chaining themselves to Court or Congress gates. They responded to the state efforts at erasing the disappeared from their lives by returning them to the public space through posting life-sized silhouettes representing their relatives.

By framing their right to know the truth in such seemingly depoliticized terms, these discursive mobilizations had the effect of making it seem natural, self-evident and unquestionable. The first bulletin of the *Madres* illustrates this idea of moral necessity, as it read: '*We don't even ask for their freedom. We are just trying to get (someone) to tell us where they are, of what they are accused, and that they be judged in accordance with the law (...) Could there be any plea more elemental, more correct, more human, more Christian?*'[20]

C. The Official Response: Truth Commissions, Exhumations, Reparations

G. Delhaye stresses that the naturalization of the norm allows for a legal innovation to stop generating controversy[21] as it gathers critical support to reach a tipping point whereby it gets broadly accepted. This stage of norm internalisation usually starts with the institutionalisation of the norm in the sense of the creation of a formal structure that accommodates the norm.

In this regard, the effectiveness of the families' mobilizations to demand truth as a moral and historical necessity is evidenced in the first national policy responses and discourse of most Latin American post-authoritarian regimes in the wake of democratic transitions from the mid-1980s into the 1990s. For instance, in Chile, in the aftermath of the Pinochet regime in 1990, President Aylwyn justified the creation of National Truth and Reconciliation Commission, an investigative body, on the grounds that '*the moral conscience of the Nation demands that the truth for the grave violations of human rights committed in our country between September 11, 1973 and March 11, 1990 be brought to light*.'[22]

The extent of norm internalisation is demonstrated by the fact that new democracies embraced the demand for truth in various ways. They ordered the exhumation of bodies from mass graves to return them to families so as to allow their decent burial; they set up 'truth commissions', namely fact-finding institutions to disclose 'truth' through official inquiries into human rights violations; some allowed selective prosecutions and implemented reparations. As an example, Argentina's National Commission on the Disappeared (*CONADEP*) documented 9,000 disappearances out of an estimated 30,000 from 1976 to 1983 and brought to light the mechanics of state

ISBN 978-0-9566139-1-2

terrorism. Its archives were then used in the 1985 trials of the military leaders. Another indicator of norm internalisation is the express reference to a right to truth of the people in the mandates of certain commissions. For instance, the mandate of Guatemala's 1994 Commission for Historical Clarification provides that '*the right of the people of Guatemala to know the whole truth regarding these events, which, if clarified will help to ensure that this sad and painful chapter will not be repeated and that the process of democratizing Guatemala will be strengthened.*'[23]

However, it soon appeared that successor regimes privileged the establishment of 'truth commissions' over nation-wide prosecutions because of political, pragmatical and legal impediments such as amnesties negotiated by outgoing regimes and aimed at shielding perpetrators or the destruction of evidence.[24] Policies aimed at 'national reconciliation' such as the 'Aylwyn doctrine' in Chile resorted to 'truth commissions' as a compromise for peace and justice, stressing the fragility of democratic transitions. In brief, violations could be investigated, but not prosecuted.[25] National policy-makers thus started to construe truth as a demand separate from justice in order to legitimate their policies. With the idea that truth commissions constituted a valid alternative to prosecutions, grassroots movements started re-claiming truth at the international level in the light of the amnesties adopted in Argentina, Chile, Uruguay, throughout the 1980's onwards, feeling that their demand for truth had been downplayed.

THE INTERNATIONAL DIFFUSION OF A RIGHT TO TRUTH

The naturalization of the demand for truth also allowed critical support for its international entry point and circulation among multilateral organizations. The resonance of the framing of the demand for truth by family-based grassroots movements vis-à-vis the international sphere is demonstrated by the language used in the UN General Assembly's resolutions since the late 1970s. For instance, it has consistently insisted on the impact of disappearances on family ties by declaring itself *'[d]eeply moved by the anguish and sorrow which such circumstances cause to the relatives of disappeared persons, especially to spouses, children and parents'*.[26]

Where one could have expected the decline of the right to truth as a mobilizing frame in reaction to its potential 'instrumentalization' at the state level by amnesties and other measures aimed at impunity, this part will show how, on the contrary, various truth entrepreneurs have gradually reframed the right to truth as part of the agenda of the fight against impunity on the international level and how this has led to the proliferation of norms among multilateral organizations to the extent that truth-seeking mechanisms have become a common denominator of peace-conflict resolution and post-conflict rebuilding practices around the world. It thus examines how the claim for truth has migrated from a national context onto the international stage through the combined effect of three factors of consolidation or legitimation in the international arena: transnational advocacy networks, the rise of a new expertise centered around the notion of transitional justice and professional migrations between the advocate, academic and political spheres.

Transnational Entrepreneurs as Catalysts for Norm Institutionalisation

In the context of a globalized world, the role of non-state actors in shaping the agenda of human rights on the international scene is being increasingly studied. Keck and Sikkink have advanced the concept of '*transnational advocacy networks*' as the main engine of transformation to explain the effectiveness of norm diffusion at the global level. It captures cooperation among NGOs as well as connections to potentially like-minded actors as in international governmental organisations, church or union organisations or the media.[27] Such networks 'shame' the human rights violator and mobilize further international support from international organizations. The above mentioned concessions by successor regimes in Latin America were evidently also the result of transnational advocacy networks. The use of expertise, exchange of information among them or placing of an issue on the agenda are all considered an important source of influence for the institutionalisation of norms by international organizations. In the case of the diffusion of the right to truth, because of the extent of disappearances in Latin America, the main organizational platforms before which transnational entrepreneurs brought their actions and denunciations have been the Organization of American States (*OAS*), especially the Inter-American Commission and Court of Human Rights, as well as the various UN agencies. The increasing visibility of these transnational norm entrepreneurs has proved instrumental for the institutionalisation of the right to truth in international organizations through the fight against impunity.

ISBN 978-0-9566139-1-2

Families as Transnational Players: Reframing Truth as an Urge to Combat Impunity

The insertion of grassroots movements into transnational advocacy networks has allowed them to advocate more effectively their right to know the truth before both local and international arenas by building direct dialogue between the international and domestic spheres while circumventing state authorities at home.

First, by reaching to international platforms, grassroots movements have become more familiar with international law and fully became human rights organizations. Estela Barnes de Carlotto, President of *Abuelas* and recipient of UN Human Rights Prize in 2003, admitted: '*When they kidnapped my daughter, I didn't know anything about Amnesty International, or the Inter-American Commission on Human Rights, or the United Nations*'.[28] Similarly, Alicia Garcia, cofounder of Salvadoran *CoMadres*, recalled that '*despite the fact that we aren't people with a lot of education, the situation obligated us to learn about national and international laws. Now there are CoMadres who cannot read or write but who can debate the socks off a lawyer about international law! And the CoMadres wins!*'[29] This integration of the international legal language changed their discourse and actions. For instance, before Western audiences, *Madres* integrated the memory of the Holocaust in their discourse and argued that if their loved ones never reappeared alive, a genocide would have been committed.[30]

Second, their engagement in transnational advocacy marks the beginning of their international campaign for a universally-binding instrument against enforced disappearances. In early 1981, *Madres* testified at the first international conference on enforced disappearances organized in France by exiled Argentine lawyers, human rights NGOs and the Paris Bar Association. That same year, 14 family-based movements throughout Latin America joined their efforts to promote a treaty against disappearances at the global level through the creation of FEDEFAM, the *Latin American Federation for the Associations of Relatives of Detained-Disappeared*, which became the largest social movement against disappearances connecting with human rights associations, Churches, unions, political parties and charity groups. FEDEFAM came up with the first draft treaty to criminalize disappearances as crimes against humanity as soon as 1982 and were instrumental to the adoption of the 1992 *UN Declaration* and 1994 *Inter-American Convention* and the 2006 *Convention against Enforced Disappearances* that expressly recognizes the right to truth.[31]

Third, the Madres and FEDEFAM had a direct influence on the integration of the right to truth on the international agenda of the fight against impunity.[32] Throughout the 1980s, they rallied 'against oblivion', and for 'memory' before the UN Commission on Human Rights in Geneva to voice that domestic amnesties and pardons prevented victims from knowing the truth and pressed for prosecutions. Their campaigns before that body sparked many initiatives and transformed the Commission into a norm entrepreneur as such for the diffusion of the right to truth. For instance, it created the Working Group on Involuntary or Enforced Disappearances in 1980, charged with assisting relatives to locate missing persons worldwide by acting as a channel of communication between families and the state. It charged UN French expert Louis Joinet to examine the issue of amnesties since 1985, which culminated in a 1997 report on the impunity of perpetrators of human rights violations. This report expressly made the right to truth a pillar and the first principle of the fight against impunity, and listed other principles such as the '*state's duty to remember*', as '*people's knowledge of the history of its oppression is part of its heritage'*, the role of extrajudicial commissions of inquiry to '*ascertain truth*' but as a *complement* to prosecutions, and the need to preserve archives of human rights violations.

International Non Governmental Organizations: Bridging Gaps Between Human Rights

Social constructivists also point to the inter-connectedness of a new right with an existing frame as a further condition likely to increase the persuasiveness of a novel claim before international platforms.[33] In this regard, landmark cases have been litigated by international NGOs (hereafter INGOs) thanks to their consultative status before international human rights bodies and their role of legal representation of victims. Because their agendas were broader than the single promotion of one norm, many INGOs such as *Amnesty International* or the *International Commission of Jurists* (*ICJ*) built creative linkages between the emergent right to truth and existing norms. For instance, the first recognition of the right to truth by the Inter-American Court on Human Rights (hereafter ICtHR) can be partly attributed to the *ICJ*'s *amicus curiae* entitled 'the right to truth' presented before the judges. Before this case, judges were reluctant to acknowledge a new 'right to truth', simply declaring that '*such a right does not exist in the American Convention'*.[34] This document was the first ground-breaking sophisticated legal argumentation that suggested that even though it was a new claim, the right to truth existed through its articulation with other existing rights such as the state duty to investigate abuse, the right to family life, or the

ISBN 978-0-9566139-1-2

prohibition on torture.[35] This thorough legal argumentation has greatly influenced the later developments of the right to truth in other judicial arenas.

The Rise of a New Expertise: "Transitional Justice" Experts as New Cause Lawyers

Links between professional mobilizations and the academic sphere are also to be explored to understand the institutionalisation of norms. Another major influential vector that has driven the institutionalisation of the right to truth concerns the self-constitution of a new professional community mobilized at the international level around the notion of 'transitional justice', a term devised in the late 1980s to refer to judicial and non-judicial mechanisms implemented to face a legacy of massive human rights violations in societies that have experienced wars or authoritarian regimes.[36] For instance, the *International Center for Transitional Justice (ICTJ)*, a New-York based institution created in 2001 with local and regional offices worldwide, regularly advises the UN and trains judges and policy-makers. Although the term refers to disparate and contingent domestic policy options such as criminal trials, truth commissions, reparations for victims and institutional reforms aimed at deterring crimes, they were assembled and rationalised on the grounds that where mass atrocities have occurred, specific measures, different from 'ordinary justice', are needed to respond to the needs of victims. Over the past decade, these new advocates, mostly human rights lawyers, have successfully imposed themselves as a new body of international expertise through two factors that are relevant to the diffusion of the right to truth at the international level.

First, these new 'experts' have become one of the most prominent 'truth entrepreneurs' or 'truth crusaders' through their active promotion of truth commissions as a generalized 'model' for peace and reconciliation. Despite the fact that these inquiry bodies were mostly working as compromises for justice at the national level, they were nonetheless 're-enchanted' by human rights advocates who had been involved in their design or implementation.[37] For instance, the co-founders of the *ICTJ* who advocate this model have all been closely involved with truth commissions. Alex Boraine was the former Vice-President of the 1994 South African Truth and Reconciliation Commission (*TRC*); Paul Van Zyl was its General Secretary; and Priscilla Hayner, the first 'specialist' of truth commissions who contributed to the elaboration of a model to rationalise them through generic definitions of truth commissions[38] and recipes for 'what works best?' among past experiences worldwide and what is 'ideal in most circumstances.'[39]

Second, the recent emergence of transitional justice as an autonomous academic discipline within universities has further consolidated the reception of truth commissions as a credible 'model' of peace and justice by international organizations, especially through the theorization of truth commissions as an institution mainly dedicated to the recognition of victims by allowing them to 'tell their stories' and the importance of 'dealing with the past'. For instance, doctrinal debates have praised them as symbols of a paradigm shift from an *accused-centered* model of *retributive justice* that focuses on the determination of individual guilt, to a *victim-centered* model of *restorative justice* by associating a larger number of victims to their proceedings and allowing evidentiary standards that are less constraining than in criminal trials. However, through the integration of transitional justice experts in universities, the literature now commonly acknowledges that truth commissions are "complementary" to criminal prosecutions.[40] The overwhelming academic enthusiasm for transitional justice is evidenced by its increasing integration in curriculums, including prestigious Law Schools.[41] Most of the *ICTJ* members teach in such universities. In 2007, Oxford publishers launched the first review entitled *'The International Journal of Transitional Justice'* specifically dedicated to the discipline... with an editorial board composed of *ICTJ* members.

The professionalisation and academic validation or endorsement of transitional justice practices, reconfigured generally as 'truth-seeking and reconciliation policies', have thus favoured their 'scientisation' by abstracting them from their context-specific and political origins. This has in turn favoured their integration into the UN agendas and their increasing application in the case of post-war scenarios.[42] For instance, truth-seeking is now invariably mobilized by UN organs to strengthen peace, reestablish the rule of law, facilitate reconciliation or reinforce the fight against impunity,[43] to the extent that it increasingly imposes the creation of a truth commission through peace agreements.[44] The idea that these politics of memory implement a nation's 'collective right to truth' or that a 'historic truth' is owed to societies that have experienced large-scale abuse is increasingly acknowledged by international and regional multilateral bodies.[45] In fact, a dynamic of norm imitation among UN agencies and regional bodies has culminated with resolutions expressly entitled 'the right to truth'[46] and UN agencies promoting 'best standards'[47] to emulate countries to implement the right to truth through truth-seeking practices and ensure that the fight against oblivion is part of the fight against impunity.

ISBN 978-0-9566139-1-2

Professional migrations between the militant, academic and political spheres

Professional migrations and trajectories between the militant, academic and political spheres are also highly relevant to understand the percolation of the right to truth in multilateral bodies and its corresponding institutionalisation.

Most 'transitional justice experts' began their careers as activists in local or international NGOs, before being involved in truth commissions or international criminal trials, and eventually becoming professors, or influential UN experts or members of human rights bodies that have recognized the right to truth. In many INGOs, the integration of the right to truth as part of these organization's agenda has also often been due to the presence of diaspora from authoritarian regimes. For instance, Juan Méndez, President of the *ICTJ* until 2009, was an Argentine political prisoner during the *junta* and has been one of the main promoters of the right to truth since the 1990s, calling it '*one of the most important issues in Latin America*'.[48] He advocated for truth during his work at *Human Rights Watch*/America's Watch[49] before migrating to the Inter-American Commission on Human Rights in 2000. His presence at that Commission coincides with the first cases brought before the Court arguing explicit violations of a right to truth and the first recognition by the Court that amnesties for gross violations violate the victims' right to truth.[50] Moreover, a closer look at the honorary members of the INGO *International Commission of Jurists*, which wrote the landmark *amicus curiae* on the right to truth before the Inter-American Court, reveals that most of them have been essential promoters of the right to truth in international law while belonging to all spheres of power. Examples include Theo Van Boven, who linked the right to truth with the right to reparations while he was a UN Special Rapporteur on restitution, compensation and rehabilitation for victims of gross violations of human rights.[51]

Professional migrations between human rights bodies and truth commissions have also helped spread the idea that a 'historic truth' is owed to societies that have experienced large-scale abuse, as prominent members of the Inter-American Commission had previously worked for truth commissions in their country of origin. To mention one emblematic example, José Zalaquett, President of the Inter-American Commission in 2004, previously sat in the *Chilean Truth and Reconciliation Commission* due to his local activism during Pinochet's regime. He had also been one of the main exporters of truth commissions abroad, serving as a legal advisor to the South African government on the establishment of the Truth and Reconciliation Commission to confront the legacy of the apartheid.

Lastly, tracing personal trajectories can also partially account for accidental processes of diffusion of an idea. For instance, the rapporteur of the first international conference that promoted a treaty against disappearances in a report entitled '*Refusing to Forget. The Politics of Enforced Disappearances,*' was… Louis Joinet, the same person who fifteeen years later authored, as a UN expert, the influential report where the victims' right to truth was framed as the first principle to combat impunity, as well as the necessary condition for societies' to avoid the repetition of violence.

CONCLUSION

This paper has attempted to map a few categories of actors, instruments and dynamics that typically remain within the margins of legal analysis when it comes to explain the 'sudden' apparition of legal innovations on the international sphere. This socio-legal itinerary has tried to demystify the process of norm creation by engaging with a few processes below the gaze of formal actors, institutions and mechanisms of international law in order to highlight some transformative micropractices that have driven the circulation of the idea of a right to truth onto the international stage. To this end, it has tried to show that the success of the right to truth is not only the result of a miraculous convergence of the international community towards a good principle,[52] but that its universality, its naturalization and resonance have been the result of deliberate social processes of construction such as framing and expertise-building through alliances with academic and political environments, pointing to the porosity of frontiers between arenas of norm creation and consolidation, and between the global and local levels. The right to truth has both served as an empowering and disempowering tool for victims' groups and advocates over time, but its endless capacity to accommodate various layers of meanings and subtexts will most likely carry it towards different journeys, as already witnessed by its present-day reappropriations by other norm entrepreneurs advocating in completely different contexts than serious violations of human rights.[53]

ISBN 978-0-9566139-1-2

REFERENCES

[1] UN General Assembly, Resolution 61/177 (2006). Article 24 (2) of the International Convention for the Protection of All Persons from Enforced Disappearances, which provides that *'(e)ach victim has the right to know the truth regarding the circumstances of the enforced disappearance, the progress and results of the investigation and the fate of the disappeared person. (…)'*

[2] UNGA, Resolution 3220 (1974), § 8.

[3] UNGA, Resolution 61/155. *Disappeared Persons.* (2006).

[4] Short, D. 'Sociological and Anthropological Approaches', in Goodhart, M. (ed.) *Human Rights. Politics and Practice*. (2009) Oxford: Oxford University Press, p. 96.

[5] Plummer, K. 'Rights Work: Constructing Lesbian, Gay and Sexual Rights in Modern Times,' in Morris, L. (ed.) *Rights: Sociological Perspectives*, (2005) London and New York: Routledge, p. 153.

[6] Finnemore, M. & Sikkink, K. (1998) 'International Norms Dynamics and Political Change,' in *International Organization*, Vol. 52, No. 4. pp. 887-917, pp. 904-905; Risse, T. & Sikkink, K. 'The Socialization of International Human Rights Norms into Domestic Practices: Introduction,' in Risse-Kappen, T., Ropp S. & Sikkink, K. *The Power of Human Rights: International Norms and Domestic Change.* (1999) New York: Cambridge University Press, pp. 16-17.

[7] Schmitz, H. & Sikkink, K. 'International Human Rights,' in Carlsnaes, W., Risse, T. & Simmons B. (eds.) *Handbook of International Relations*. (2009) London: SAGE, 2nd ed., p. 523.

[8] Scovazzi, T. & Citroni, G. *The Struggle against Enforced Disappearances and the 2007 United Nations Convention*. (2007) Leiden-Boston: Martinus Nijhoff Publishers, p. 8.

[9] *Le refus de l'oubli. La politique de disparitions forcées. Colloque de Paris janvier-février 1981*, (1982) Paris: Berger-Levrault, p. 230. (Intervention by the Mothers of the Plaza de Mayo. The translations are mine.)

[10] Lievesley, G. *Democracy in Latin America. Mobilization, power and the search for a new politics.* (1999) New York: Manchester University Press, p. 133.

[11] Finnemore, M. & Sikkink, K., *op. cit.*, p. 897. See also Snow, D. (2004). 'Framing Processes, Ideology and Discursive Fields.' in Snow, D., Soule, S., Kriesi, H. (eds). *The Blackwell Companion to Social Movements*. Oxford: Blackwell Publishing, pp. 382-412.

[12] Boli, J. & Thomas, G. (1999) 'INGOS and the Organization of World Culture,' in Boli, J. & Thomas, G. (eds.) *Constructing World Culture: INGOS Since 1875*. Stanford: Stanford University Press.

[13] Waters, M. (1996) 'Human Rights and the Universalisation of Interests,' *Sociology*, Vol. 30, No. 3, 593-600, p. 593.

[14] Bennett, J. & Simpson, J. *The Disappeared and the Mothers of the Plaza.* (1985) New York: St Martin's Press, p. 160.

[15] Westwood, S. & Radcliffe, S. 'Gender, Racism and the Politics of Identities in Latin America,' in *'ViVa'. Women and Popular Protest in Latin America.* (1993) Londres-New York: Routledge, p. 18.

[16] Lievesley, G., *op. cit.*, p. 136; Fisher, J. *Out of the Shadows: Women, Resistance and Politics in South America.* (1993) London: Latin America Bureau, p. 135.

[17] 14th March of Resistance of Madres (Linea Fundadora), available at Madres' website: http://www.madres.org/marchas/resistencia.asp

[18] Brysk, A. *The politics of human rights in Argentina. Protest, Change, and Democratization* (1994) Stanford: Stanford University Press, p. 48. See also Franco, J. (1987) 'Gender, Death and Resistance; Facing the Ethical Vacuum,' *Chicago Review*, Vol. 35, No. 4, pp. 59-79.

[19] *Le refus de l'oubli. La politique de disparitions forcées*, *op. cit.*, p. 230.

[20] Madres de Plaza de Mayo, *Boletín Informativo*, cited in Ricardo Leís, H. *El movimiento por los derechos humanos y la política argentina*. (1989) Buenos Aires: Centro editor de América Latina, p. 100.

[21] Delhaye, G. 'Comprendre la mondialisation des normes : les leçons d'échec de la liberté religieuse universelle,' in Schemeil, Y. & Eberwein, W. (ed.), *Normer le monde.* (2009) Paris: L'Harmattan, p. 93.

[22] Supreme Decree No. 355 of April 25, 1990, Report of the *Comisión Nacional de Verdad y Reconciliación* (National Truth and Reconciliation Commission), Volume I.

[23] Agreement on the establishment of a 'Commission for Historical Clarification of human rights violations... that have caused suffering to the Guatemalan people,' preamble, § 2, in Kritz, N. (ed.) *Transitional Justice. How Emerging Democracies Reckon with Former Regimes*, vol. III. (1995) Washington DC: United States Institute of Peace, p. 220.

[24] Pinto, M. *L'Amérique latine et le traitement des violations massives des droits de l'homme.* (2007) Paris: Pedone, p. 26.

[25] Cardenas, S. *Human Rights in Latin America: A Politics of Terror and Hope.* (2009) Philadelphia: University of Pennsylvania Press, p. 176.

[26] UNGA Res. 33/173 (1978), preamble, § 5. See also, for instance, Res. 45/165 (1990), preamble, § 3.

[27] Keck M. & Sikkink, K. *Activists Beyond Borders: Advocacy Networks in International Politics.* (1998) Ithaca-London: Cornell University Press.

[28] *Ibid.*, p. 93.

ISBN 978-0-9566139-1-2

[29] Schirmer, J. 'The Seeking of Truth and the Gendering of Consciousness: The CoMadres of El Salvador and the CONAVIGUA widows of Guatemala,' in *'ViVa', op. cit.*, p. 30.
[30] *Le refus de l'oubli*…, *op. cit.*, p. 230.
[31] Scovazzi, T. & Citroni, G. *The Struggle against Enforced Disappearances and the 2007 United Nations Convention*. (2007) Leiden-Boston: Martinus Nijhoff Publishers, p. 98.
[32] Joinet, L. (1997) *Final report on the question of the impunity of perpetrators of human rights violations*…, Commission on Human Rights, E/CN.4/Sub.2/1997/20/Rev.1, p. 3.
[33] Finnemore, M. & Sikkink, K., *op. cit.*, 938.
[34] *Inter-American Court of Human Rights*, Castillo Páez v. Peru case, merits, 3 Nov. 1997, series C, n° 120.
[35] ICJ. (2001) 'Legal Brief Amicus Curiae Before the Inter-American Court of Human Rights in the Case of Efraín Bámaca Velásquez v. Guatemala', in *Review of the International Commission of Jurists*, No. 62-63, pp. 127-158.
[36] Bickford, L., « Transitional Justice », in Shelton, D. (ed.), *The Encyclopedia of Genocide and Crimes Against Humanity*, *Vol. 3*. (2004) Detroit: Macmillan, p. 1045.
[37] On this specific question, see Lefranc, S. (2009) 'La professionnalisation d'un militantisme réformateur du droit: l'invention de la justice transitionnelle,' in *Droit & Société*, No. 73, pp. 561-590, p. 567; Lefranc, S. (2008) 'La justice transitionnelle n'est pas un concept,' *Mouvements*, 2008/1, No. 53, pp. 61-69.
[38] 'Truth Commissions', entry by Hayner, P. in Shelton D. (ed.) *The Encyclopedia, op. cit.*, pp. 1045-1047.
[39] Hayner, P. *Unspeakable Truths. Confronting State Terror and Atrocity*. (2001) New York: Routledge, pp. 321-322.
[40] Roht-Arriaza, N. & Mariezcurrena, J. (eds) *Transitional Justice in the Twenty-First Century: Beyond Truth versus Justice*. (2006) Cambridge: Cambridge University Press.
[41] To mention a few examples, a specific class on transitional justice is offered at the Law Faculties of Columbia, NYU, Essex University, as well as in many political sciences faculties such as French *Institut d'Etudes Politiques*.
[42] See Commission on Human Rights, Resolution 2005/70 entitled 'Transitional justice'; Jouhanneau, C. (2008) "Si vous avez un problème que vous ne voulez pas régler, créez une Commission'. Les commissions d'enquête locales dans la Bosnie-Herzégovine d'après-guerre,' in *Mouvements des idées et des luttes*, No. 53, Paris: La Découverte.
[43] See *for eg.* the following UN Security Council Resolutions: Res. 1888 (2009); Res. 1757 (2007); Res. 1606 (2005).
[44] For instance, the international community has imposed to parties to peace agreements the obligation to set up truth commissions in Liberia, Sierra Leone, Guatemala, Salvador, Burundi and the DRC.
[45] UN Secretary-General Report on the 'Rule of law and transitional justice in conflict and post-conflict societies.' (2004) S/2004/616, §§ 50-51; Human Rights Council (2008), A/HRC/9/L.23, § 2; Report of the Parliamentary Assembly of the Council of Europe (2007). 'Use of the experiences of 'truth commissions.'
[46] Commission on Human Rights, Res. 2005/66; Human Rights Council, Res. 9/11 (2008); Parliamentary Assembly of the Council of Europe, Res. 1463 (2005); OAS, GA Resolutions 2175 (2006), 2267 (2007) and 2406 (2008).
[47] UN High Commissioner for Human Rights, *Rule-of-Law Tools for Post-Conflict States. Truth Commissions*, New York, Geneva: UN, 2006, p. 27 and its 'Study on the Right to Truth,' E/CN.4/2006/91.
[48] Antkowiak, T. (2001) 'Truth as Right and Remedy in International Human Rights Experience,' *Michigan Journal of International Law*, Vol. 23, p. 980.
[49] See for instance HRW/America's Watch, 'Truth and Partial Justice in Argentina,' (1991) http://www.hrw.org/en/reports/1991/04/01/truth-partial-justice-argentina.
[50] Inter-American Court of human rights, *Castillo-Páez*, merits. *(1997)* and *Barrios Altos case v. Peru*, merits (2001).
[51] In 'Study on restitution, compensation and rehabilitation for victims of gross violations of human rights,' E/CN.4/Sub.2/1993/8, general recommendation n° 14, p. 54, § 134, 137 (11).
[52] See Lefranc, S., 'La production de nouvelles techniques de pacification : la normalisation internationale de causes locales,' in Schemeil, Y. & Eberwein, W. (eds). *Normer le monde.* (2009) Paris: L'Harmattan, pp. 65-86.
[53] See for instance Article IXX's coalition campaign for Global Transparency Initiative claiming a 'right to know' in the context of the accountability of international financial institutions, at www.ifitransparency.org. In the United States, the creation of a 'Financial Truth Commission' has also been suggested by certain senators to investigate further than the mandate of the official commission of inquiry into the subprimes crisis (*New York Times*, 29 July 2009).

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

A CASE FOR INTERDISCIPLINARY STUDIES: IN PRAISE OF THE POLYFOCAL LENS

Lyndal NEELIN

PhD Student, Carleton University, Ottawa, Ontario, Canada
lneelin@connect.carleton.ca

Keywords: Multidisciplinary, Interdisciplinary, Poly-focal, Margins, Resilience

ABSTRACT

My dissertation research is an interdisciplinary project, supported by an interdisciplinary school, supervised by a multidisciplinary committee. This paper attempts to articulate what these terms mean to me as a (re)searcher, and to justify my choice to situate myself and my project between the lines of disciplined research. To do this, the paper clarifies the distinction between the terms inter- and multi-disciplinarity, and attempts to reclaim the margins and peripheries of academe for un-disciplined research. I will argue (with Foucault) that the margins of disciplinary orthodoxy constitute unique creative spaces precisely because they are uncomfortable.

My case study site, the village of Shawville, Québec, is perched uncomfortably in multiple border spaces. It sits on the line between French and English Canada, between urban centre and hinterland, between productive and re-creative space. I use tools from anthropology, political economy, history and geography to examine the mixture of ideational and material factors that contribute to the community's resilience.

Together, the disciplines create a poly-focal lens that draws a kaleidoscope of factors into focus: Shawville's social, political, economic and geographic locations. Inter-disciplinary space allows the kaleidoscope lens to turn, breaking and making connections, shuffling and re-weighting the factors at play. As Charles Tilly argues, when we move from one theoretical stance (or one discipline) to another the whole frame of reference changes, it becomes not only a question of competing explanations, but of what 'counts' as evidence, what needs to be explained, and even how an explanation must be structured in order to be heard[1]. Between frames, the questions, answers, and what counts as 'knowing' participate in the discussion rather than containing it.

INTRODUCTION

> *"That may be true, he acknowledged, but it's completely accurate, and as long as the answer is right, who cares if the question is wrong"*[2]

The case for interdisciplinary scholarship that this paper develops rests on the conviction that questions *do* matter, that they are, in fact, something we should care about more than we seem to. The questions we ask, like the stories we tell, become the tools of our collective imagination.[3] Questions limit and shape the topics that get examined. They speak volumes about the assumptions and values of individual researchers. They reflect the worldview and concerns that drive the agendas of research agencies and funding organizations. In the quote above, the speaker contends that a 'right' (i.e. accurate and provable) answer is all that matters. This paper will argue that without the 'right' (i.e. relevant and incisive) question it is quite possible for an answer to be altogether meaningless. At its best, interdisciplinary scholarship incorporates the forming and framing of questions as part of the research problem.

ISBN 978-0-9566139-1-2

My recent experience defending my dissertation proposal exposed a very particular constraint on the sort of questions expected of academic research. In order for my project to be approved I had to demonstrate that the questions I wanted to ask were questions that had answers. I was asked to outline specifically how I planned to look for those answers, how I would recognize them when I found them, and how I would communicate them in my finished work. Posing questions that could be answered by following a thoughtfully constructed plan rendered my project 'doable' – high praise indeed. The problem with this process, the sharpening of my original, admittedly 'fuzzy', questions into finely tuned research instruments, is that it pushed us, my supervisory committee and me, to think backwards, to begin with the answers we anticipated and think through the data that would prove salient to arrive at the questions that would uncover it. The process privileged, by default, our preconceptions. I could not ask a question whose answer I could not imagine.

Thinking backwards troubles me on two counts. The first is that finely honed instruments cannot be expected to find things they are not looking for. The second is the idea that information is a *thing* that can be found. Even as the social sciences acknowledge in theory the insight that our 'reality' is an inter-subjective product, in practice we have yet to shake the ideal of objective data that can be collected and analyzed. When we allow our questions to be shaped by anticipated answers we risk reinforcing and reproducing existing constructs. It is hard to get to a new place following a tried trail.

In the context of my project, my hope is that this limitation is offset by the supervision of a multi-disciplinary committee. Each of them is looking at a different map. Their *super*vision, watching my progress from different perspectives and with different expectations will ensure that there is no straight and certain path. I am almost sure to get lost for a while in the wilderness between their expectations and, if I can weather the dust storm, learn something no one is expecting.

My dissertation project explores the resilience of a small Ottawa Valley town (resilience defined here as the capacity to absorb disturbance without changing form)[4]. The town is interesting because it has held onto a distinct character despite strong odds against it. Shawville is among the last English-speaking majority communities in Canada's Francophone province of Québec. In the interviews I have been conducting, community members have described the town as 'tight' (close knit, closed and quick to close ranks), a place with a long memory and deep family ties. Fifteen hundred and eighty seven people live in the town. Statistics Canada 2006[5] describes the community as predominantly English-speaking (eighty-five percent report speaking English 'most often at home'), remarkably homogeneous (the census reports a total 'visible minority' population of ten), and unusually sedentary, especially given the well-documented predisposition of English-speaking Quebeckers to move on to greener pastures since the rise of the Québec separatist movement in the mid 1960s.[6] Two thirds of Shawville's residents have roots in Shawville for three or more generations. The numbers in Statistics Canada's community report reflect Shawville's aging population, the 'dip' in the number of people between twenty and forty years old that is characteristic of English Québec, and the strangely small number of people employed in 'occupations unique to primary industry' given the area's designation as a 'resource region' and its long-time identification with agriculture and forestry.

Shawville's location *in* Québec, and the concomitant political and economic pressure to speak French as the '*shared public language*'[7] of Québec, makes its enduring 'Englishness' conspicuous and therefore intriguing. It suggests that 'Englishness' is in some way a defining characteristic of the community, even as it points to the potential costs of insisting on a trait which is arguably maladaptive within the community's socio-political environment. How is it that this community endures? What defines it? What supports and sustains it? How has it managed the social, political and economic turbulence of recent decades to maintain its sense of self and keep its balance? My work explores the context for, and the implications of, Shawville's persistence as an English-speaking, rural, and homogeneous community within the Francophone, urban-focused and increasingly diverse province of Québec. In exploring Shawville's resilience, I hope to learn something about the strengths and needs of rural, agricultural communities in Canada more generally.

It is easy to jump to several conclusions about Shawville. In the context of Québec's struggle for recognition as a distinct society, Shawville is a conservative Anglophone hold-out. In Canada's increasingly urban and multi-ethnic reality, Shawville is a white, rural backwater. As the drivers of Canada's economy have moved west to Alberta's oil fields, Shawville has been left behind. But Shawville is also a community rich in natural beauty,

ISBN 978-0-9566139-1-2

history, and relationships. Its resilience is supported by some alchemy between its location, its economic base, and its socio-linguistic identity. The value of approaching a complex 'problem' like Shawville's resilience from multiple angles is obvious. Juxtaposing the disciplinary predispositions of history, geography, political economy, and linguistic anthropology (the disciplines represented by my multi-disciplinary supervisory committee) already goes some way toward mitigating the problem of thinking backwards from a presumed scenario to the questions worth asking. Allowing the variety of possible angles to actually destabilize each other, beginning to think in the spaces and contests between disciplines, opens up the possibility of different questions altogether. Is it possible, for example, that being 'left behind' economically has been good for Shawville, increasing its resiliency by insulating the community from external dependence, and forcing local economic diversification? Has its 'tight' garrison mentality[8] played a role in this resilience, perhaps by serving to mitigate the risks in temporary or seasonal unemployment? Has the lack of mobility (and the strong pull of family) slowed the drain of people and resources experienced by all rural communities sufficiently to let this community keep its feet?[9]

Arjun Appadurai, writing about the experience of 'modernity', argues that our reality is so complex and interconnected that once we begin to ask questions we find that each one leads to another. The varied and overlapping layers of modern experience are infinitely 'nested'. Appadurai reminds us that wherever we choose to begin, whatever heuristic point we adopt, '*the last turtle is always a matter of methodological convenience or stamina*'.[10] It is important to remember that what we 'know' is contingent, fluid, and contestable.

Liora Salter and Alison Hearn define interdisciplinarity as '*any challenge to the limitations or premises of the prevailing organization of knowledge*', concluding that, '*It is the act of challenging what would otherwise be taken for granted as the proper organization, content, methodology, or purpose of research that creates and defines interdisciplinarity*'.[11] I did not set out to challenge the taken-for-granted organization of knowledge. I simply failed to take it for granted. One by-product of moving between continents as a child is the certainty that today's 'normal' will be tomorrow's 'weird'. There is almost nothing that cannot be done or conceived of just as readily in an altogether different way. Once you begin to think like that, an epistemological discussion is already engaged.

METHODOLOGY

In much of the literature the term *inter*-disciplinary is used to describe any research project that blurs or crosses disciplinary lines. Julie Thompson Klein, writing about interdisciplinary studies, uses the same term to mean '*both instrumental borrowing across disciplines and the development of new conceptual categories*'.[12] Salter and Hearn, co-editors of *Outside the Lines: Issues in Interdisciplinary Research,* also identify two distinct 'camps' in the interdisciplinary field: one espouses '*an applied or problem-centered view of knowledge*'[13] while the other is engaged in '*a theoretical, primarily epistemological enterprise*'.[14] These are, in my mind, two very different projects. I find it helpful to distinguish between them by calling practical, applied research 'multi-disciplinary' and reserving the term 'inter-disciplinary' for conceptual or epistemological work. Joe Moran, a scholar of interdisciplinarity, also distinguishes between multidisciplinary work as '*the simple juxtaposition of two or more disciplines*' and interdisciplinary work which is '*always transformative in some way*'.[15] Perhaps the distinction is not entirely fair since any movement between established categories of thought and practice is bound to open up some discussion about what 'counts' as evidence, as an explanation, or as an answer. As soon as you begin to raise those sorts of questions you are pushing at accepted ideas about knowledge and knowing. The distinction I am pointing to is that while some researchers look for practical solutions to bridge these gaps, others seize on the gaps to pry open orthodoxies and politicize knowledge.

ISBN 978-0-9566139-1-2

Figure 1: Conceptualizing Shawville

In order to conceptualize Shawville (Fig. 1) my project steps onto territory traditionally associated with geography, political economy, history and anthropology, borrowing conceptual categories from each of these disciplines in order to 'thicken'[16] the representation of my case study. In this sense my project is multi-disciplinary. I have adopted research tools and methods from a variety of disciplines in order to answer a research question that crosses disciplinary lines. My research tool kit includes archival research, the editorial pages of the weekly local paper between 1970 and the present, interviews with key community figures, and a bi-weekly column in the local paper in which I outline my recent 'discoveries' and invite comments from those who remember the events and upheavals I describe. I am looking for clues to the question 'What sustains this community?' in the tools and strategies people have used to mitigate and cope with destabilizing changes. I focus on events because while 'change' is often too nebulous to be examined, specific incidents associated with change can be: events like the closing of the Hilton Mine in 1977, the separatist referenda in 1980 and '95, and the 'mad cow' (BSE) scare in the early 2000s. In many ways, *how* an event has been remembered is as significant as *that* it has been remembered. Alessandro Portelli, an historian, writes that memory of an event itself '*is only a key to everything that happened before and after it… [so that] errors, inventions and myths lead us through and beyond facts to their meanings'*.[17] Michael Agar, an anthropologist who describes himself as a 'professional stranger', describes the rewards of managing to step outside our own assumptions and listen to what we are being told:

> *If you just stop asking people questions framed in terms of the way you already see things, if you just listen and struggle with what patterns and passions drive their talk, you get a glimmer of a different kind of life. And that glimmer, brief and fragmentary though it might have been, hints that what you thought you knew was way off base*[18].

These "rich points", as Agar calls them, are moments that allow us to discover assumptions we didn't know we had made, and other assumptions we have never considered.[19]

ISBN 978-0-9566139-1-2

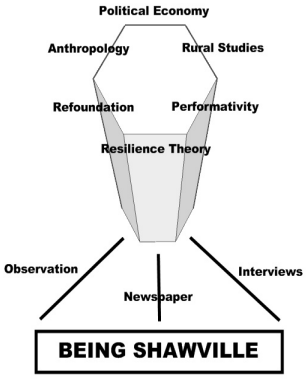

Figure 2: Shawville under the kaleidoscope

The goal of my interviews and archival reading in Shawville is to bring the near view into focus, to give me a sense of the breadth and depth of the ground under the community's feet. It is also important to keep Shawville's context in view, and look for connections to events unfolding at the provincial, national, and larger scales. I am counting on the poly-focal (multi-disciplinary) lens I am using (Fig. 2) to correct for both myopia and presbyopia, making it possible to see both distant happenings and what is directly in front of my nose. I rely on the spaces between the disciplinary perspectives to allow for movement, and the contests between those perspectives to create it. Here I find the image of a kaleidoscope helpful. Microscopes and telescopes are static. They are best suited to examining a single slide or snapshot view, one slice of time. A kaleidoscope is designed for motion. Multiple mirrors and a rotating lens shuffle bits of cotton and glass into infinitely varied arrangements, rearranging them anew with every twist. A kaleidoscope offers a way to catch a glimpse of the multiple, varied and overlapping layers of Appadurai's '*modernity*'[20], in the same way that clouds moving quickly overhead will sometimes make us conscious of the ground moving beneath our feet. It is difficult to keep the near view, the long view, and the whole view-in-motion in focus. As Appadurai grieves, we have no theory '*sufficiently quirky*'[21] to help us hold onto the necessary level of complexity. Fortunately, because theory and method are tied together, we can use analytic tools and methods to push our conceptual abilities. *How* we do research will shape what we see, and what we see will drive the questions that we learn to ask.

DISCUSSION

Discipline…

Multi-disciplinary…

Inter-disciplinary…

Undisciplined…

Perhaps the best way to enter into a discussion of disciplined/undisciplined research is to follow Fredrik Barth's example and look to the boundaries around and between disciplines to ask what purpose the divisions serve, and how they are maintained. In the introduction to *Ethnic Groups and Boundaries* Barth challenged anthropology to shift its focus onto '*the ethnic* boundary *that defines the group, [away from] the cultural stuff that it encloses*'.[22] His insight was that the 'why' and 'how' of ethnic identity could be more informative than the 'what' of an ethnicity's identifying traits. Applied to an examination of academic disciplines, Barth's insight suggests that we focus, not on the '*constellation of methods, perspectives and topics*'[23] that characterize any one discipline, but on the protocols and gatekeepers that divide them.

Joe Moran describes the development of disciplines as '*an organic consequence of advances in knowledge, [in combination with] institutional and societal factors, particularly the demand for specialists in a complex and technologically sophisticated society*'.[24] He challenges the ready association of disciplined knowledge with industrialization by reminding us that '*anxieties about the harmful specialization of knowledge*' are already evident in Aristotle's concession that '*the ordering of knowledge*

ISBN 978-0-9566139-1-2

into disciplines was necessary but regrettable'.[25] Salter and Hearn trace the development of disciplines back to the medieval university's division of theology and art from medicine and law, an act they attribute to external pressure from the associated professional guilds. They distinguish between these early categories and the '*true disciplinarity and specialization*' of the nineteenth century, which they link to the '*rise of the modern natural sciences…and the specific demands [for specialization] of the industrial marketplace*'.[26]

The question underlying the contest between an ancient origin for disciplinarity and the suggestion that it is a relatively 'new' innovation is whether the divisions are inherent in the nature of knowledge and learning, or the product of external institutional, ideological or market forces. I am not convinced that this is the 'right' question. Although it clearly makes a difference whether our understanding of knowledge is constrained by its own logic or by the logic of 'the market', the nature of the constraints themselves suggest a more fundamental question about knowledge and power. It is instructive, for example, that one of the current arguments *against* interdisciplinary studies picks up the theme to grieve that changes in the university structure, this time *away* from traditional disciplines, are being driven not by intellectual, but by political and financial concerns.[27] Stated this way, the distinction implies that as long as the changes/norms/constraints are intellectually driven, all is well. While it is generally agreed that allowing a political agenda, or free market logic to drive policy within our universities is not desirable, it seems naïve to assume the benevolence of an intellectual agenda. Perhaps our efforts are better spent thinking about the effects of these constraints on knowledge, rather than struggling to ascribe their origins to either set of interests.

In *The Discourse on Language*, Michel Foucault describes how traditional disciplines (and their attendant discourses) successfully shape ideas, perceptions and expectations through various 'rules of exclusion'; prohibitions drawn around some topics, some situations, and some people so that not just anyone can say anything, anywhere.[28] These prohibitions are enforced through rules of expected form and acceptable authorship. According to Foucault the boundaries between disciplines create internal constraints around what can be said and who can say it, by establishing research agendas and recognizing hierarchies of status and experience.[29] Religious, juridical and academic rituals give some people and some utterances credibility while their absence denies credibility to others. Disciplinary structures and 'fellowships of discourse' limit access to certain knowledge, and defensive fences enclose orthodoxy, keeping some ideas inside, and all others out.[30] Bruce Knauft, calling on Foucault to explain constraints on knowledge in the context of his own work, calls this effect '*the envelope of the unthought and the unsaid*'.[31] Foucault's rather sinister description of disciplined knowledge underpins a critique of the way disciplines work to shape particular subjectivities, and limit access to certain knowledges.

Another effect of disciplinary divisions has been to dislocate and fragment knowledge, making room for its reification, and limiting opportunities to make important connections across fields. Donald Wilson, a Canadian historian, likens the increasingly narrow focus of specialization in the academy to working on a fence line by digging the post holes ever deeper without returning to nail the connecting rails between the posts.[32] Both Klein and Moran, describing the advent of the modern university, point to the expectation that individual specialization would operate within a community of learning, so that a collective grasp of connections and a holistic view of knowledge would not be lost.[33] The barriers to interdisciplinary work that Salter and Hearn describe, suggest that the integrity of this 'community of sharing' has been compromised. It is not possible to believe that issues like the commensurability of concepts and terminology, or the question of audience receptivity apply to exchanges between disciplines only when they are called interdisciplinary.

A further effect of disciplines has been to create a hierarchy of knowledge, in which a Western 'scientific' worldview is conspicuously privileged. Accepting 'science' as the sole metric of worth has meant accepting a Eurocentric measure of what constitutes knowledge, and so who holds the 'real stuff' and who does not. Science insists that questions have identifiable answers. Definitions must be operationalized; the success of any intellectual inquiry must be measurable and verifiable. All the social sciences have struggled with the need to measure themselves against '*the language of the exact sciences [as] a kind of absolute*'.[34] When scholars protest that interdisciplinary work runs the risk of becoming 'undisciplined' or 'facile' they frame their concern in terms of lost standards for measuring excellence and/or the validity of research findings[35], and they accuse interdisciplinary schools of being '*haven[s] for weak and fuzzy thinking*'.[36] I believe that the posturing, the complaints and accusations, have two objectives. The first is to affirm the 'central' position of the traditional disciplines by

ISBN 978-0-9566139-1-2

pointing to their appropriate measures of excellence and legitimacy. The second is to close ranks against any challenge to the epistemology that supports the Westernized status quo.

Gayatri Spivak, an eminent post-colonial theorist, argues that the same '*programming function*' of discourse that disciplines knowledge and silences dissenting questions also marginalizes the subaltern 'other'.[37] She points to '*a whole set of knowledges that have been dis-qualified as inadequate or insufficiently elaborated: naïve knowledges, located low down on the hierarchy, beneath the required level of cognition or scientificity*'.[38] The bearers of such knowledges, those who live and know within such epistemes have no ground to speak from within the Western science-centric parameters of what 'counts' as knowledge. In order to be heard they must borrow acceptable metrics and vocabularies, in effect, relinquishing their distinctness. They cannot speak unless they cease to be themselves.[39]

Working to establish the credibility of feminist epistemology, Dorothy Smith also addresses the issue of marginalization. She describes her own experience of exclusion and invisibility within the academy. Smith contends that social scientific methods *'created a standpoint from which the reader reflected on [his or] her life as if [he or] she stood outside it'*, disqualifying as 'not knowledge' '*their own actual situations and…their own good knowledge of the practicalities and organization of their everyday and everynight worlds*'.[40] Smith suggests that the problem is much deeper than the absence of women as researchers and/or subjects. Her sense of exclusion lay in the way the '*presuppositions of [social scientific]discourse…denied [her] presence*'.[41] Judith Butler's often referenced concept of 'performativity' suggests that woman-ness is partially constituted in that exclusion.[42] If it is true that identity is 'tenuously constituted' in the performance of ritualized actions, then it can be changed by interrupting the performance; a 'legacy of sedimented acts' can still shift and resettle.[43] I read a caution in Butler's observation that when you make something that was invisible visible, you are not in fact looking at the once-invisible thing, what you see has always already been transformed. When you examine the 'once-invisible-thing-made-visible' you are examining a new 'visible' thing.[44] In her critique of main stream social science Smith challenges her discipline to make room for women, not by asking them to change their experience and perspectives, but by altering the assumptions underpinning the endeavour of social science. Obviously, alternative positionalities are only valuable in the pursuit of understanding if they remain alternative.

I have referred specifically to the marginal position of subaltern knowledges and feminist epistemes. Obviously, I could go on to discuss other ways to interpret the world which are excluded from mainstream thought within the academy[45]. My objective is not to grieve their exclusion, but to celebrate their existence, and to note that they develop and thrive as alternative worldviews precisely because they are marginalized. Foucault observed that 'disempowerment *provides unique spaces of creative marginality'*.[46] If this is true, empowerment within the system would put the creative spaces occupied by interdisciplinary studies at risk. The possibility that interdisciplinarity might be too easily co-opted by the promise of a more comfortable status within the academy and policy-making institutions more generally, is something to guard against. That said, Foucault's own description of the '*rules of exclusion*'[47] makes it very clear that establishing credibility and/or being heard from the margins will pose a challenge. There is plenty of evidence within the literature on interdisciplinary studies to suggest that life on the margins of disciplinary orthodoxy is neither comfortable nor prestigious. If the possibility of holding onto an alternative perspective, or of thinking new thoughts, depends on locating ourselves on the margins, interdisciplinary scholars might have to accept disempowerment within the establishment as part of the bargain.

While Salter and Hearn's reference to undisciplined research above was pejorative, Moran promotes undisciplined, even "promiscuous" research, as constructive and essential. While acknowledging that the potential for gaps and errors that worries the critics of interdisciplinarity are real, he insists that *'if a certain messiness goes with the territory, this is also what makes the territory worth occupying*'.[48] Interdisciplinarity has been described as '*a state of unpreparedness*' [on the part of researchers], a readiness to step beyond certainty onto unfamiliar and even contested ground.[49] In stark contrast to this deliberate vulnerability, Salter and Hearn report that '*members of the academy [within disciplinary folds] feel they are adequately prepared…to determine the proper organization of knowledge'*.[50] (1997:157) If this is true it begs for a discomfiting push outside the boundaries. Work between disciplines is important because it keeps the lines of communication (between disciplines, between epistemes, between researcher and subject) open, the translation of concepts and terms in process, the use of language alive, and avenues for reception open – all work that can be done only from the margins.

ISBN 978-0-9566139-1-2

CONCLUSION

Merrill Singer, an anthropologist, claims that the central question of all social scientific inquiry is '*What power, what interests, wrap this local world so tight that it feels like the natural order of things to its inhabitants?*'.[51] My research project explores this naturalized-order-of-things as it is experienced in the community of Shawville, Québec. My goal is to unwrap and de-naturalize Shawville's experience in order to examine the impacts and implications of the status quo on Shawville's present and future capacity for resilience.

The kaleidoscopic lens I have chosen as a way to conceptualize Shawville helps me to keep the community's past and present, as well as its near and distant contexts, in focus and in motion. Shawville is perched in a linguistic contact zone; how much of its experience is reflected in the literature about border zones, or sociolinguistic analysis? The community has shown itself to be determinedly English-speaking; how might it's 'Englishness' be understood in relation to its Protestant Irish identity? Is it significant that the families that established Shawville left Ireland before the famine, and were well established in their new homes before the Catholic Irish arrived in Canada in great numbers? Are there echoes of their chain–migration settlement patterns in the choices today's teenagers make about where (and whether) to study after high school? What is it that brings so many of Shawville's sons and daughters 'home' for their retirement after a lifetime working and raising a family 'away'? Under the kaleidoscope, patterns are formed and broken. Connections I had imagined important give way to connections I had not thought to make.

What counts as a question remains central to my project. The careful rendering of a research proposal that fulfilled the requirement for operationalized questions and a clear agenda felt like poring over a recipe before checking the cupboards to see what ingredients were on hand. Hearing from individuals within the community has given me the chance to poke about in the pantry. By paying attention to 'what' gets remembered and 'how' it is remembered I am beginning to gain a sense of what categories might count, what questions might be worth asking.

Engaged in exploring a local community as if it were a foreign place, I am often reminded of an experience I had some years ago at a workshop preparing volunteers to go overseas. At one point, while we were discussing the challenges of learning a new language, one of the other workshop participants exclaimed in relief that she 'already had that covered'. Where she was going, she would be able to hit the ground running and be immediately useful. Deep in our own fears of finding ourselves useless, we were all impressed. Not the workshop leader; he was quick with his warning that without the enforced silence of several months without words there was a very real danger of doing harm. He assured us that it would take no longer to learn the local language than to recognize local categories of thought and patterns of interaction. It would serve us well to watch for a while. Words alone would not tell us what counted as a problem or a solution. It would take us the most time of all to figure out what questions needed asking.

REFERENCES

1 Tilly, Charles. *Identities, boundaries and social ties*. (2005) Boulder, Colo.: Paradigm Publishers, p. 30

2 Juster, Norton. *The Phantom Tollbooth*. (1995) London: Collins, p. 175.

3 Daniel Francis claims that, "The stories we tell about the past produce the images we use to describe ourselves as a community. If we are not telling ourselves the right stories [i.e. stories that adequately reflect who we are, and explain to us where we are], then we cannot imagine ourselves acting together to resolve our problems" *National Dreams: myth, memory, and national history*. (1997) Arsenal Pulp Press. p.176. I want to suggest that questions have a similar significance in both reflecting where we are, and in determining future directions.

4 Gunderson, Lance. Ecological Resilience – In Theory and Application. *Annual Review of Ecology and Systematics*. 31 (2000) pp 427-8.

5 Shawville, Quebec (table). 2006 Community Profiles. 2006 Census. Statistics Canada Catalogue no. 92-591-XWE.Ottawa.http://www12.statcan.ca/census-recensement/2006/dp-pd/prof/92-91/index.cfm?Lang=E Released March 13, 2007. (accessed November 29, 2009).

6 Rudin, Ronald *The forgotten Quebecers: a history of English-speaking Quebec, 1759-1980*. (1985) Québec: Institut québécois de recherché sur la culture: Distribution, Diffusion Prologue, pp. 29,30,89,163.

7 Oakes, Leigh and Jane Warren (eds.) *Language, Citizenship and Identity in Québec*. (2007) Basingstoke [England]; New York: Palgrave Macmillan, p.84.

ISBN 978-0-9566139-1-2

[8] Northrop Frye famously describes Canadians as having a 'garrison mentality', living in communities that focus inward, with their collars up and their backs to the cold.
[9] Antonio Gramsci wrote that "the rate of change is often no less important than the direction of the change itself"– citing the example of the British 'enclosure movement' and the way in which the landed gentry's opposition to the industrialist's expropriation of lands bought time for the dispossessed peasants to adapt to their 'freedom' to work for wages. *The great transformation: the political and economic origins of our time / Karl Polany***;** foreword by Joseph E. Stiglitz ; introduction by Fred Block. Boston. (2001) MA.: Beacon Press, 2nd Beacon Paperback ed, p. 39.
[10] Appadurai, Arjun. *Modernity at Large: Cultural Dimensions of Globalization*. (1996) Minneapolis: University of Minneapolis Press. p.11.
[11] Salter, Liora and Alison Hearn (eds.). *Outside the Lines: Issues in Interdisciplinary Research*. (1996) Montreal: McGill-Queen's University Press, p.43.
[12] Klein, Julie Thompson. *Interdisciplinarity: History, Theory, and Practice*. (1990) Detroit: Wayne State University Press, p. 27.
[13] Salter and Hearn op.cit., p.8.
[14] ibid. p.9.
[15] Moran, , Joe. *Interdisciplinarity*. (2002) London and New York:Routledge, p.16.
[16] Geertz, Clifford. "Thick Description" pp. 1-30 in C. Geertz *The Interpretation of Cultures*. (1973) New York: Basic Books.
[17] Portelli, Alessandro. "Introduction" and "The Death of Luigi Trastulli: Memory and the Event", in *The Death of Luigi Trastulli, and Other Stories: Form and Meaning in Oral History*. (1991) Albany, N.Y.: State University of New York Press, p.1-2.
[18] Agar, Michael. *The Professional Stranger*. (1996) San Diego: Academic Press, p.38.
[19] ibid. pp. 33-34.
[20] Appadurai describes the overlapping layers of modernity as fractal and polythetic. op.cit. p. 46.
[21] ibid. p. 33.
[22] Barth, F. "Introduction" Pp. 9-38 in *Ethnic Groups and Boundaries* F. Barth ed. (1969) (Results of a symposium held at the University of Bergen, 23rd to 26th February 1967.) Bergen, Universitetsforlaget; London: Allen and Unwin, p.15.
[23] Salter and Hearn op.cit. p. 22.
[24] Moran op. cit. p.13.
[25] ibid. p. 5.
[26] Salter and Hearn op.cit. p. 19.
[27] Klein, op.cit. p.35.
[28] Foucault, Michel. "The Discourse on Language" Pp. 215-237 in Michel Foucault "*The Archeology of Knowledge and the discourse on Language*" (1970) New York: Pantheon, p.216.
[29] ibid. p.222.
[30] ibid. pp. 225, 226.
[31] Knauft B. "Foucault Meets South New Guinea: Knowledge, Power, Sexuality" *Ethos* 22 (1994) p.391.
[32] Donald Wilson in Berger, Carl. (ed.) *Contemporary approaches to Canadian history*. Toronto: Copp Clark Pitman, (1987) p. 232.
[33] Klein op.cit. p. 20, and Moran op.cit. p. 5.
[34] Salter and Hearn op.cit. p, 14.
[35] ibid. p. 33.
[36] ibid. p. 85.
[37] Spivak, Gayatri Chakravorty. "Can the Subaltern Speak?" pp. 271-313 in Cary Nelson and Lawrence Grossberg (eds.). *Marxism and the Interpretation of Culture* (1988) London: MacMillan, p.282-3.
[38] ibid. p. 281.
[39] ibid. p.285.
[40] Smith, Dorothy. "Feminist Reflections on Political Economy" *Studies in Political Economy* 30 (1989) p. 39.
[41] ibid. p. 52.
[42] Butler, Judith. "Performative Acts and Gender Constitution: An Essay in Phenomenology and Feminist Theory" (1988) *Theatre Journal* 40(4) p. 523.
[43] ibid. p. 519.
[44] ibid. p. 523.
[45] See Robyn Eckersley (eco-centric environmentalism), Bruno Latour ('earthly' science), Ferber and Nelson (post-autistic economics) inter alia.
[46] Knauft op. cit. p. 415. (emphasis mine)
[47] Foucault. op. cit. p.216.
[48] Moran op.cit. p. 184.
[49] Salter and Hearn op.cit. p. 157.
[50] ibid. p. 157.
[51] Singer, Merrill. Community-Centred Praxis: Toward an Alternative Non-Dominative Applied Anthropology *Human Organization*. 53 #4 (1994) p.26.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

MEN DOING 'WOMEN'S WORK': Male Domestic Workers in Nigeria

Zahrah NESBITT-AHMED

PhD Student, London School of Economics and Political Science, London, U.K.
Z.Nesbitt-Ahmed@lse.ac.uk

Keywords: Domestic Work, Gender, Men, Masculinities, Gender division of labour

ABSTRACT

Men and women are increasingly moving into gender atypical jobs and while there is extensive literature on women's increased labour force participation, labour market segmentation and segregation and the effect this has on gender relations and family decisions, there is relatively little research on men in general and men who perform what could be seen as 'women's work' in particular. The tendency to overlook issues concerning men may reflect gender studies dominating focus on women and the absence, until recently, of issues concerning men and masculinities. Very few studies deal with gender differences in domestic work – mainly because domestic work is seen entirely as a woman's job. Domestic work is a form of employment that is very interesting to analyse men doing 'women's' work because it disputes the notion that care work is reserved only for women. Through in-depth interviews with 19 male and female domestic workers, 5 employers and 2 key informants in Lagos, Nigeria this research delves into the lives of domestic workers to gain an insight into their gendered experiences. In domestic work, men and women are ascribed different roles – women are often found in work related to the domestic sphere (carers and cleaners) while men are found both inside (carers and cleaners) and outside (drivers, security guards, gardeners) the domestic sphere. Also, while male domestic workers might engage in higher paying domestic work, females might engage in unpaid domestic work the employers' home. By including men in the analysis, it is hoped to introduce a male perspective in a field of research that is predominantly female and present a broader understanding of domestic work. This paper does not intend to dismiss the strong gender divisions of labour or issues of sexism and gender inequality but can help complicate our understanding of the idea and process of 'gendering' in domestic work.

INTRODUCTION

Theoretical work on gender and employment is wide-ranging and has been concerned with women's participation in the workforce and, as Chant with Craske highlight, 'their relegations to less skilled and less remunerated jobs'.[1] These perspectives have focused on the question of female labour force participation labour market segmentation and segregation, the effect it has on gender relations and family decisions and what Simpson terms 'token women'.[2] Considerable attention has been paid to women but not so much has been paid to men when it comes to issues of gender divisions of labour. As Chant and Gutmann state, the limited studies of men in employment can be linked more broadly to the fact that within development, 'gender is still largely equated with women alone'.[3] Although there are good reasons why women should have been the focal point of attention, Chant with Craske note that 'the implied acceptance of men's economic activity as a given also reinforces the idea that employment and breadwinning constitute fixed and universal components of male behaviour and identities'.[4]

Writings on men and masculinity, 'men in crisis', 'troubled masculinities' and 'men at risk',[5] reflect what Cornwall see as 'a belated recognition that men also have gendered identities'.[6] These studies enable spaces to be opened up for reflection about how men can be disempowered or marginalised. Searlee-Chatterjee, for example, shows

ISBN 978-0-9566139-1-2

that men of the Sweeper caste in Benares hide themselves from public view when they cook or clean pots, as men who have to cook for themselves are mocked in many societies.[7]

As highlighted by Meagher, 'economic restructuring has precipitated changes in the gender division of labour within informal labour markets across Africa'.[8] While women are more pronounced in the informal economy than men, and heavily concentrated in domestic labour, petty trade, and local crafts, 'in the face of high unemployment and intense competition within the informal economy, men have begun to shift into more lucrative women's informal opportunities, such as women's tailoring and informal cosmetics manufacturing in West Africa, and beer brewing in East Africa'.[9] Overa claims that other 'female' niches men are going into are 'trade in second-hand clothing and African and African souvenirs. A few men also appear among the increasing numbers of hair-braiders'.[10]

It cannot be disputed that domestic work has traditionally, and continues to be, conceptualised almost exclusively as 'women's work'. This has led to the particular themes and concepts in the domestic work literature addressing female domestic workers. Yet this deceptively simple picture needs to be complicated and research on domestic work needs to examine the role of male domestic workers. The experiences of male domestic workers is something that is required because it shows that care work does not have to be restricted to women and that it is not 'naturally' their domain. In doing so, socially constructed gender roles and expectations can be confronted and challenged. This paper will look at the experiences of male domestic workers in Lagos.

DOMESTIC WORK

Domestic work is one of the oldest and most important occupations for millions around the world.[11] Yet, data on the number of domestic workers throughout the world are hard to collect. Available data from the ILO shows that domestic work absorbs a significant proportion of the workforce: 'in developing countries it accounts for between 4 and 10 per cent of total employment (both female and male), compared to industrialised countries where the figure ranges between 1 and 2.5 per cent of total employment'.[12]

In the majority of studies on domestic work, the worker has been female and this has then been taken to be definitive and the most significant feature of domestic service.[13] As Ebery and Preston put it for England, between the seventeenth century and the Victorian period 'service was redefined and transformed from a largely male, often undifferentiated role, to an occupation which was female dominated, highly differentiated and almost entirely domestic'.[14] It is this social transformation and cultural redefinition that led Katzman, Salmon, Dudden, Van Raaphorst and others studying domestic work in nineteenth and twentieth-century USA and Western Europe,[15] to view, as Moya puts it 'the association of women and domestic service as a permanent, almost natural, condition'.[16]

According to Bujra, these studies argue that the dominance of women in domestic work simply reflects the fact that paid domestic work is 'an extension of female domestic duties in the home' ... as women are transferring skills learnt at home to the marketplace'.[17] Furthermore, women are preferred domestic works purely because of their docility and the belief that they are more used to having to obey.

Clearly female labour is central to domestic work, but what is striking in these studies is the absence of men despite the fact that male domestic workers are present in many parts of this sector. This, according to Kilkey and Perrons, means that 'very little is known about male domestic workers and the types of work they do'.[18] Historical studies on domestic work in Europe and North America point to a 'masculinisation' of domestic work. For example, Faye Dudden notes that although domestic work was considered a female occupation in nineteenth-century United States, one-third of adult domestic workers were black men.[19] This was mainly because of the discriminatory nature of the labour market where blacks were excluded from nearly all other types of work except domestic service, thereby pushing Black (as well as Chinese) males into 'an otherwise female occupation'. Farley Grubb has shown that over two-thirds of the servants who emigrated from England and Ireland to colonial America in the seventeenth century were young men.[20] And Galenson, also on writing on America, discussed that they continued to be mostly male during the first half of the eighteenth century.[21] Trends of 'masculinisation' of domestic work can be found in Renaissance Venice where Dennis Romano found in the

ISBN 978-0-9566139-1-2

16th century that 'an aristocratic style of servant keeping with an emphasis on male servants replaced a bourgeois or mercantile style more concerned with female servants that was typical of the 14th ad 15th century'.[22]

Recent studies are pointing to a 're-masculinisation' of domestic work.[23] Even in Latin America, where Moya and others, note that 'domestic service is an overwhelmingly feminine activity',[24] Thomson notes that men are 'employed but usually as gardeners, watchmen, or chauffeurs'.[25] In India, Mehta shows that men predominated domestic work form much of the twentieth century; while in Tanzania and Zambia, Bujra and Hansen, respectively, show that the domestic work force is predominantly male.

These studies highlight that the 'feminisation' of domestic work is not fixed and as Moya puts it 'although domestic service is generally identified nowadays as a female sector it was not always so, and it is still not so in some parts of the world'.[26] As Bujra notes in Tanzania, where men are employed to clean and cook, to wash and iron, to garden and occasionally even to mind babies, 'men's successful performance of domestic service ... is neither "natural" nor a product of their usual socialisation … [and] it is clear that an apprenticeship in gender subordination is not essential to acquiring domestic skills for wage work'.[27]

Therefore, in order to gain a broader understanding of domestic work and the gender dimensions of domestic work there is a need to include in the literature the analysis of male domestic workers'. According to Manasalan, including male domestic workers' 'does *not* in any way disavow the strong gender factor in labour market segmentation nor dismiss issues of sexism and gender inequality … [but] can help nuance and complicate our understanding of the idea and process of 'gendering' in the domestic industry'.[28] Placing the experiences of men alongside that of women highlights the continuities and discontinuities of domestic work, complicates the idea of care work and prevents us from falling into the normative and universalising trap of implicitly regarding women as natural nurturers (ibid).

Domestic Work in Nigeria

Under the ILO's International Standard Classification of Occupations (ISCO) a domestic worker is 'someone who carries out household work in private households in return for wages'.[29] The current labour law of Nigeria does not define domestic workers. Therefore domestic workers' in this paper are persons who are recruited from outside the employing household, and paid by wage or 'in kind' to perform labour in and around the household. Domestic workers in this study include 'house-helps', drivers, 'maiguards', cooks, washer-men, gardeners and nannies.[30] They could either be live-in or live-out, part-time or full-time, skilled or unskilled.

To date, there is limited data on the number of persons employed as domestic workers in private households in Nigeria. There can be little doubt to anyone familiar with Nigeria that domestic work is a part of Nigerian society - Nigerian families have always had access to domestic help and share a culture that supports the practice of having a family or non-family member perform domestic labour. Domestic work is also an important source of employment for many of the urban poor from rural areas and neighbouring countries such as Benin, Togo, and Ghana.

Domestic work in Nigeria did not constitute wage employment until during the colonial period. In the colonial times, British households usually included adult domestics such as stewards, cooks, and drivers, as well as young domestics known as 'small boys' and 'baby nurses'. During this period, only a few elite Nigerians (as well as repatriates from Sierra Leone and Cuba) in the then capital, Lagos, emulated the British tradition to employ domestics. Others obtained the assistance of their older and younger relatives, who in exchange for domestic chores and chid-care would receive educational or vocational training. This was known as fostering.[31]

Child-fostering is a common practice in West Africa. Guinean child domestic workers, for example, often work in the house of a relative, where they have been sent by their parents at an age as young as five.[32] Children serve as house-helps or shop assistants on the understanding that their foster families will help them to pursue diverse educational activities and/or pay monthly sums of money to parents (or children) as payment for their children's labour. As Assan notes in his study in Accra, employers commonly see themselves as the kind saviours of children who would be destitute without the food and shelter they provide.[33] Unfortunately, employers fail to make the promised education or payments to either the children or their parents. Furthermore, the fostering institution has undergone much change. In the past it used to be done mainly through direct contact between the

ISBN 978-0-9566139-1-2

child's parents and employers; however, this link with the parents and sense of obligations has increasingly been eroded, as more and more children are procured from impoverished rural families by middlemen to work in urban households.

Poverty is the greatest single force that creates the flow of people into domestic work for their own and for their families' survival. Another factor, as highlighted by Okafor, is the growing entry of women into the labour market in the modern Nigerian society.[34] In a traditional Nigerian society, males were accorded more value and prestige; while women derived their social status from their roles as a wife and mother. With increased access to education, gender awareness and urbanisation, most of the traditional barriers and stereotypes hindering the advancement of women have been broken down. As more women enter the labour market, the pressure and challenge of maintaining cherished traditional values of home keeping, has led to the rise in domestic workers as 'a substitute labour to assume the woman's role in the household'. In this way, the woman-employer is freed from her household responsibilities and, in turn, is able to take up paid employment or other activities outside the home (ibid).

METHODOLOGY

The study was qualitative in nature and included questionnaires and in-depth interviews (IDIs) with domestic workers, employers and key informants. The study was conducted in Lagos, South-western Nigeria, between July and August 2009 because as Rukoo notes 'domestic work is mainly concentrated in urban areas, with a minority rural petty bourgeoisie employing some domestic workers'.[35] Lagos is a socio-cultural melting pot with diversity of employers and domestic workers making it an ideal site for the research.

Since domestic workers are hidden behind closed doors, snowball sampling was applied in the selection of respondents as the most convenient and suitable sampling methodology. Domestic workers and employers were approached via Nigerian friends and acquaintances; and the purpose of the study explained to them. The identified domestic workers and employers then introduced the researcher to other domestic workers and employers.

Questionnaires were utilised to get background information of all domestic workers. The questionnaires covered questions related to socio-demographic characteristics, occupation, salary, and living and working conditions. Once questionnaires were administered, employers had a say on allowing their domestic workers take part in IDIs to help gain deeper insight into the living and working conditions of domestic workers, the challenges they face, and their relationships with employers. A total of 19 domestic workers (8 female and 11 male), between the ages of 16 and 55, were reached. It was hoped to conduct IDIs in a public place because Esim and Smith notes that this 'ensure[s], as much as possible, the freedom to respond truthfully'[36] but with the exception of 2 female domestic workers, interviews happened in the workplace (i.e. the employers home). To complement the selection of domestic workers, 5 employers were selected. Specifically, the research wanted to know why employers employed domestic workers, their gender preferences and recruitment method. In addition, interviews were done with 2 key informants. Interviews with key informants were selected on the basis of their experience and knowledge of domestic workers. A total of 26 interviews were held. Finally, the data collected were analysed using largely qualitative methods. Limited quantitative methods were used to present the socio-demographic data of respondents using percentages. Throughout the period of the study the consent of the domestic workers, employers and key informants were sought and obtained so as not to violate respondents' free will and privacy. In any case, where anyone declined or objected to be included in the study, another respondent in another household was selected.

NOT JUST 'WOMEN'S' WORK: MALE DOMESTIC WORKERS IN LAGOS

Agadjanian notes that men's presence in occupations that are (or are seen by most people as) predominantly female reflects changes in the urban labour market.[37] In his study on male street vendors in Maputo, Agadjanian notes that men's movement into 'women's' occupations is precipitated by socioeconomic downturns and the accompanying reduction of employment opportunities.[38] Another important factor, as noted by Overa in his

ISBN 978-0-9566139-1-2

study on men in the informal economy in Accra is that the majority of the men who enter 'female' domains in the informal economy have migrated, and therefore perform their new roles out of sight of kin, mates and elders of their home towns.[39]

In Lagos, 7 of the men interviewed chose domestic workers because of lack of alternative employment while 4 chose it because of low pay at previous employment. Majority of males in domestic work view it as a means to an end. Statements such as 'money has no face', 'everyone must chop', 'family responsibilities', shows that even if men may feel embarrassed engaging in domestic work, the severity of the situation in the labour market justifies their choice. Furthermore, male domestic workers are usually hoping or working towards better job prospects. For example, a driver discussed how he was saving up money from his job to buy a taxi and become his own boss.

Male domestic workers are preferred as drivers, guards, cooks, washer-men and gardeners in urban households. This is because in Lagos driving, cooking, washing/ironing clothes and gardening are seen as requiring skills and specialization which women do not possess. For example, women are not deemed appropriate washer-men because they are not able to handle heavy items and are not good at ironing. They also have more defined tasks. In households that have drivers, the drivers washed and polished the cars, and drove members of the household to and fro different destinations. Some households also employ washer-men whose job is to wash and iron clothes. Security guards were employed to guard the house.[40]
The only form of domestic employment both men and women are found in is house-helps. As put by one female employer, those households which employed male house-helps deemed them 'hardworking and more reliable than the female ones in terms of house chores'.[41] Furthermore, employers prefer men because it is thought that they can learn more rapidly than females and their performances, when trained, are of a higher standard. Men are also seen as 'more efficient with normal housework', and 'more useful, because they will work both inside and outside'. Girls, on the other hand 'can't do heavy work' and 'they are slower than boys and they like to chatter'.[42] This shows, as Bujra put it, 'that the skills and attitudes appropriate to domestic work are a product of the structure of relationships (power, hierarchy, solidarity) in the workplace itself, rather than an outcome of processes of gender socialisation, especially when the material base of such socialisation is marked by class (and often ethnic) distinctions'.[43]

Male domestic workers share similar problems to female domestic workers. An example is their invisibility which derives primarily from the characteristics of the workplace itself - the private household and what Bakan and Stasiulis state is the 'low status of unpaid household work which is performed by women'.[44] Other major problems encountered at the workplace, include long hours of work, heavy workloads, limiting rest and leisure time, inadequate accommodation and food (live-in workers), lack of privacy and interference in personal matters, job insecurity, low salaries, absence of benefits normally granted to other categories of workers, and exposure to violence and abuse.

Despite these similarities, male female domestic workers share different experiences in domestic work. For example, a marked feature of domestic work noted by Anderson (2001) was that significant numbers of workers had accommodation with their employers.[45] In Lagos, male domestic workers are generally more successful in negotiating private living arrangements away from their place of work – of the 11 domestic workers interviewed only 3 lived in with their employer. Male live-in domestic workers are less common because as stated by a female employer some 'families do not feel safe having men sleeping in their homes, especially if they have young daughters'. Those families that do have live-in male domestic workers are able to do so because they have separate living accommodation for them in their homes, known as the boys-quarter.[46] According to King, living arrangements 'can clearly enhance or hinder employers' control over [them]'.[47] Being live-out male domestic workers are able to have some degree of autonomy from employers on a daily basis, this is something most female domestic workers do not have as they are usually live- in and 'always on call, making their lives outside the employer's home nonexistent'.[48]

Another difference observed is in the hours of work and free time. While it is hard to determine actual working times of domestic workers, it can be ascertained that the average working hours per day for female domestic workers was 14.5 as compared to 12 for male domestic workers. Male domestic workers had more fixed working hours than female domestic workers and this could be due to the fact they were live-out. Furthermore, being

ISBN 978-0-9566139-1-2

mainly live-outs gave male domestic workers the further advantage of days off – this was usually Saturday evening to Sunday afternoon or evening off.

Men also earned more than women with the highest paid female domestic worker receiving NGN16000 per month compared to NGN25000 for drivers.[49] The lowest paid male domestic workers are house-helps who receive NGN7000 per month. Overall, average female monthly salaries were NGN6777, compared with NGN15375 for males. The higher salaries of male domestic workers are due to the fact that men are seen as doing jobs that require skills and as such they are paid more for doing these 'skilled-jobs'. Although male house-helps earned the least out of all the male domestic workers because the work they did 'typically entails unpaid labour traditionally performed in the household by women, thereby making it undervalued in monetary terms'[50]; they still earned more than female house-helps- reflecting gender biases in earnings even within the same occupation. This is because male house-helps are still seen as possessing specialised skills, such as being able to wash heavier clothes and iron- things women are deemed not capable of doing. As stated earlier, males mainly come into domestic work because they have no alternatives, and while some men indicated they would like to save and leave the occupation, they said they were unable to because their money mainly went to taking care of their family. Again this could be another reason why males earn more money than females' because asides from having 'specialised' skills they are seen as the major breadwinners and so are paid more to take care of their family. Furthermore, male domestic workers seemed to have better non-remunerative benefits as they were more likely to receive housing allowance, transport money and medical care.

CONCLUSION

In including men in an analysis of domestic work, it has been shown that research on men's employment and in particular men doing women's work may have profound implications for gender identity and inequality as men doing 'women's' work tend to reduce gender barriers. For example, men and women are employed to housework, often doing similar jobs such as cleaning and caring and work long hours in the day. Men and women domestic workers are also seen as part of the same social and occupational group. They are also exposed to the similar living and working conditions of long working hours and abuse (although women face more sexual abuse). This represents, to some extent, an erosion of gender segregation within informal labour markets but this does not eliminate gender inequality.

Gender differences are still present in the nature and type of work done – men employed as drivers, gardeners, washer-men, security guards or house-helps and women employed as house-helps or nannies. In general, men are more likely to live-out and work shorter hours, with more money and days off. There are also what Agadjanian calls 'symbolic differences'[51]. This refers to the fact that women and men are ascribed different characteristics. For example, women are seen as docile traits typical of a 'woman's work style' while men are seen as having 'specialised' skills.

So while gender segregation is not entirely diminished, the presence of men and women in atypical jobs in the domestic work does represent erosion of gender segregation domestic labour markets. As Agadjanian puts it, it is,

> *the segment of the urban population least integrated into the modern urban economy and therefore most economically disadvantaged [that] may approach gender equality at a faster pace than the rest of society- exactly because its poverty and social vulnerability deprive men of important leverage of gender discrimination.*[52]

Domestic work, which occupies both men and women in Lagos shows that domestic work does not have to be restricted to women and that it is not 'naturally' their domain.[53]

REFERENCES

ISBN 978-0-9566139-1-2

[1] Chant, Sylvia with Craske, Nikki. *Gender in Latin America* (2003)Latin American Bureau, p.195
[2] See for example, Chant, Sylvia with Craske, Nikki. *Gender in Latin America* (2003) Latin American Bureau; Chant, Sylvia and McIlwaine, Cathy. *Geographies of Development in the 21st Century* (2009) Cheltenham: Edward Elgar; Chen, Marty 'Informality, Poverty and Gender: Evidence from the Global South' (2010)and Heintz, James 'Women's Employment, Economic Risk and Poverty' (2010) in Sylvia Chant (ed.) '*The International Handbook of Gender and Poverty'* Cheltenham: Edward Elgar; Simpson, Ruth 'Gender, Mix and Organisational Fit: How Gender Imbalance at Different Levels of the Organisation Impacts on Women Managers', (2000) *Women in Management Review*, p.350
[3] Chant, Sylvia and Gutmann, Matthew C. (2002) 'Men-streaming' gender? Questions for gender and development policy in the twenty-first century', *Progress in Development Studies* 2:4, p.271
[4] Chant, Sylvia with Craske, Nikki. *Gender in Latin America* (2003)Latin American Bureau, p.195
[5] Chant, Sylvia and Gutmann, Matthew C. (2002) 'Men-streaming' gender? Questions for gender and development policy in the twenty-first century', *Progress in Development Studies* 2:4,p.271-2
[6] Cornwall, Andrea (1997) 'Men, masculinity and 'gender in development', *Gender and Development*, 5:2, p.11).
[7] Searle-Chatterjee, M.(1979), 'The Polluted Identity of Work: A Study of Benares Sweepers', in S. Allman (ed.), *Social Anthropology of Work,* ASA Monograph 19, 269-85 (London: Academic Press)
[8] Meagher, Kate (2010) 'The Empowerment Trap: Gender, Poverty and the Informal Economy in sub-Saharan Africa' in Sylvia Chant (ed.) '*The International Handbook of Gender and Poverty',* (Cheltenham: Edward Elgar)p.475
[9] ibid; see also Chant, Sylvia and Pedwell, Carolyn (2008) '*Women, Gender and the Informal Economy: An Assessment of ILO Research and Suggested Ways Forward*' (International Labour Office, Geneva)
[10] Overa°, Ragnhild (2007) 'When Men Do Women's Work: Structural Adjustment, Unemployment and Changing Gender Relations in the Informal Economy of Accra, Ghana', *Journal of Modern African Studies*, 45:4, p.558.
[11] International Labour Organization (2010) '*Decent Work for Domestic Workers*, *International Labour Conference, 99th Session'*, p.1.
[12] ibid.
[13] These studies include Anderson, Bridget (2000) '*Doing the Dirty Work? The Global Politics of Domestic Labour'* (London: Zed Books); Cock, Jacklyn (1989) '*Maids and Madams: Domestic Workers under Apartheid*' (The Woman's Press Limited); King, Alison (2006) '*Domestic Service in Post–Apartheid South Africa: Deference and Disdain'* (Ashgate: Aldershot)
[14] Ebery, Mark and Preston, Brian (1976) '*Domestic Service in Late Victorian and Edwardian England 1871-1914*', Reading Geographical Papers, Reading University, Department of Geography, p2.
[15] Dudden, Faye (1983) '*Serving Women: Household Service in Nineteenth Century America'* (Wesleyan University Press); Katzman, David (1978) *'Seven Days a Week: Women and Domestic Service in Industrializing America'* (Oxford University Press); Salmon, Lucy Maynard (1972) '*Domestic Service'* (New York: Arno Press); van Raaphorst, Donna L. (1988) '*Union Maids Not Wanted: Organising domestic Workers, 1870-1940'* (New York:Praeger
[16] Moya, Jose C. (2007) 'Domestic Service in a Global Perspective: Gender, Migration, and Ethnic Niches', *Journal of Ethnic and Migration Studies*, 33: 4, p.565.
[17] Bujra, Janet (2000) '*Serving Class: Masculinity and Feminisation of Domestic Service in Tanzania'* (Edinburgh University Press).
[18] Kilkey, Majella and Perrons, Dianne (2010) 'Gendered Divisions in Domestic-Work Time: The Rise of the (Migrant) Handyman Phenomenon', *Time and Society*, 19:1, p13.
[19] Dudden, Faye (1983) '*Serving Women: Household Service in Nineteenth Century America'* (Wesleyan University Press), p222, 327.
[2020] Grubb, Farley (1992) 'Fatherless and Friendless: Factors Influencing the Flow of English Emigrant Servants*', The Journal of Economic History*, 52:1, p.90.
[21] Galenson, D.W. (1978) 'British servants and the Colonial Indenture System in the Eighteenth Century', *The Journal of Southern History*, 44:1, p.43-4.
[22] Romano, Dennis (1991) 'The regulation of domestic service in Renaissance Venice', *Sixteenth Century Journal*, 22:4,
[23] These studies include Kilkey, Majella and Perrons, Dianne (2010) 'Gendered Divisions in Domestic-Work Time: The Rise of the (Migrant) Handyman Phenomenon', *Time and Society*, 19:1, 1-26; Ramirez, H and Hondagneu-Sotelo, Pierette (2009) 'Mexican Immigrant Gardeners: Entrepreneurs or Exploited Workers?', *Social Problems,* 56: 70–88; Sarti, R. (2006) 'Domestic Service: Past and Present in Southern and Northern Europe', *Gender and History* 18: 222–45.
[24] Moya, Jose C. (2007) 'Domestic Service in a Global Perspective: Gender, Migration, and Ethnic Niches', *Journal of Ethnic and Migration Studies*, 33: 4, p.562.
[25] Thomson, Marilyn (2009) 'Workers not maids- organising household workers in Mexico'. *Gender and Development*, 17:2, p.282.
[26] Moya, Jose C. (2007) 'Domestic Service in a Global Perspective: Gender, Migration, and Ethnic Niches', *Journal of Ethnic and Migration Studies*, 33: 4, p.560.
[27] Bujra, Janet (2000) '*Serving Class: Masculinity and Feminisation of Domestic Service in Tanzania'* (Edinburgh University Press), p.73.
[28] Manalansan, Martin. F (2006) 'Queer Intersections: Sexuality and Gender in Migration Studies', *International Migration Review*, 40:1, p.239.
[29] International Labour Organization (1990) '*International standard classification of occupations'* (Geneva)

ISBN 978-0-9566139-1-2

[30] This is a note. 'House-helps' is the term commonly used in Nigeria to refer to workers involved with cleaning and caring tasks. They could either be 'house-girls' or 'house-boys'. 'Maiguards' is the name commonly used to refer to household security guards.
[31] This is a note. Fostering is the assumption of rights and duties of parenthood by adults who are not the child's natural parents without the latter surrendering their full rights.
[32] Human Rights Watch (2007) '*Bottom of the Ladder: Exploitation and Abuse of Girl Domestic Workers in Guinea'*, p.4.
[33] Assan, Joseph Kweku (2008) 'Generational Differences in Internal Migration: Derelict Economies, Exploitative Employment and Livelihood Discontent', *International Development and Planning Review*, 30:4, 377-398.
[34] Okafor, Emeka Emmanual (2009) 'The Use of Adolescents as Domestic Servants in Ibadan, Nigeria*', Journal of Adolescent Research*, 24:2,p.189.
[35] Rukoo (20005) cited in Platform for Labour Action (2007) '*Adult Domestic Workers in Uganda: An Analysis of Human Rights and Social Injustice'*(Fountain Publishers: Uganda), p.14
[36] Esim, Simel and Smith, Monica (2004) '*Gender and Migration in Arab States: The Case of Domestic Workers'* (ILO: Geneva),p.26
[37] Agadjanian, Victor (2005) 'Men Doing "Women's Work": Masculinity and Gender Relations Among Street Vendors in Maputo, Mozambique' in Lahoucine Ouzgane and Robert Morrell (eds.) in *African Masculinities: Men in Africa from the Late Nineteenth Century to the Present*, (Palgrave Macmillan), p.262.
[38] Ibid,
[39] Overa°, Ragnhild (2007) 'When Men Do Women's Work: Structural Adjustment, Unemployment and Changing Gender Relations in the Informal Economy of Accra, Ghana', *Journal of Modern African Studies*, 45:4, p.559.
[40] Nesbitt-Ahmed, Zahrah (2009) *Domestic Service in Nigeria: A Gendered Perspective*. Unpublished MSc in Urbanisation and Development dissertation, Department of Geography and Environment, London School of Economics, p.20
[41] Mrs. S.O-Employer
[42] Field notes
[43] Bujra, Janet (2000) '*Serving Class: Masculinity and Feminisation of Domestic Service in Tanzania'* (Edinburgh University Press), p.73.
[44] Bakan, Abigail B. and Stasiulis, Daiva (1997) 'Introduction' in Abigail B. Bakan and Daiva Stasiulis (eds.) '*Not one of the Family: Foreign Domestic Workers in Canada*' (University of Toronto Press),p.10.
[45] Anderson, Bridget (2001) '*Just Another Job? Paying for Domestic Work'*, *Gender and Development*, 9:1, 25-33
[46] This is a note. This was invented during colonialism to represent the servants quarters in the colonial masters homes.
[47] King, Alison (2006) '*Domestic Service in Post–Apartheid South Africa: Deference and Disdain'* (Ashgate: Aldershot),p.52).
[48] ibid
[49] This is a note. NGN250 is GPB1
[50] International Labour Organization (2010) '*Decent Work for Domestic Workers*, *International Labour Conference, 99th Session'*
[51] Agadjanian, Victor (2005) 'Men Doing "Women's Work": Masculinity and Gender Relations Among Street Vendors in Maputo, Mozambique' in Lahoucine Ouzgane and Robert Morrell (eds.) in *African Masculinities: Men in Africa from the Late Nineteenth Century to the Present*, 257-270 (Palgrave Macmillan),p.264
[52] ibid
[53] This is a work in progress. Please do not cite or circulate without the author's permission

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

THE PHYSICAL SPACE OF A PROBATION OFFICE: CONTROL, RISK AND PUNISHMENT

Jake PHILLIPS

PhD Student, University of Cambridge, United Kingdom
jrp60@cam.ac.uk

Keywords: Probation, Architecture, Control, Risk, Punishment, Presentation of self

ABSTRACT

By drawing on the fields of criminology, architecture, education and sociology this paper shows how the physicality of a probation office can be considered both integral to, and representative of, several important changes in the Probation Service's recent history. Probation can be seen to be on the margins of criminal justice in that it receives less financial attention from policy makers, especially in light of recent budget cuts, 'prisoncentricity'[1] and a lack of awareness of probation amongst members of the public. Research in the fields of education and public health has explored the impact of different building design on the choices and achievements of staff and pupils, whereas criminological research has tended to look at architecture in the context of prisons with a focus on two themes: that a prison building can be used to control its users, and using the building as a lens through which to look at changing attitudes to crime and punishment.

This paper brings these ideas together to explore how the architecture of a probation office creates and perpetuates a particular way of working with offenders through analysis of ethnographic work conducted at a probation office. I suggest that the relationship between the 'protected' zone of the office and the 'unprotected' zone of the waiting area and interview rooms creates a 'frontstage' and 'backstage' of probation work.[2] I examine how different parts of the building can be seen as representative of the power held by a probation officer and the increasing hegemony of risk management. I then consider how the managerialist-inspired policy of 'office consolidation'[3] represents a removal of probation from the community it is supposed to serve, as well as being representative of the reformulation of probation as a punishment in the 1990s.

INTRODUCTION

This paper addresses a question about the Probation Service which has received little attention: what is the architecture of probation; and how does this architecture affect the work of probation and reflect recent changes in probation policy? It is important to consider the Probation Service because, despite being responsible for more offenders than the Prison Service,[4] it receives considerably less attention from policy makers, politicians and sentencers in an era that has been described as 'prisoncentric'.[5] Looking at architecture in general is a useful and informative exercise because, as McAra and Armstrong explain, '*the places where anger management groups meet, the rooms where parole hearings are held, … and especially … the administrative spaces of offices*' where workers and managers create reports, Initial Sentence Plans, Pre-sentence reports and risk assessments are '*equally important [as the "monumental imagery of the supermax"] and define the nature of the penal experience. They are the places where penality takes shape, containing and conferring meaning on the objects that are necessary to translate policy into practice*'.[6] Examining the architecture of probation is important because, as will be seen in this paper, changes in policy have been reflected in the design of the probation office, which is then reflected in aspects of practice. This is all the more compelling because attention has recently been (re)directed towards the importance of the offender-officer relationship, desistance theory and offender engagement suggesting that the re-emergence of rehabilitation[7] may be here to stay. This puts us in a position where probation offices reflect a policy focussed on risk, punishment and control whereas policy itself is moving towards a more rehabilitative and engaged method of working with

ISBN 978-0-9566139-1-2

offenders. This paper argues, therefore, that if policy does continue to move towards a more rehabilitative way of working, policy makers need to consider the impact of offices which were designed for something altogether different because buildings *do* have an impact on the way people behave and work.

The paper begins with a discussion of the way criminologists have traditionally looked at architecture. As per Armstrong and McAra, much emphasis is placed on the external imagery of the prison as opposed to a micro-sociological account of how space shapes penality. That said, some criminologists have addressed this idea in the context of the prison but not probation. I then turn to the field of education, not because this provides us with the best work on architecture but to illustrate some ways in which space is seen to impact on behaviour. The empirical aspect of the paper follows a brief methodological introduction. I argue that the split between generic and specialist offender management work is evident in the design of buildings and that this impacts on the way offender managers perceive victims workers, to take one example. I also argue that the protected aspect of the office creates a zone of total confidentiality which allows offender managers to speak freely about offenders although this could, at times, be perceived as 'too free'. I then introduce Goffman's work on the 'presentation of self' because the office area and the area where offenders are seen can be conceptualised as the front and back-stages of probation. However, the reception area holds a backstage difficulty for offender managers which results in the building holding disciplinary power of both offenders and offender managers. I also argue that the physical layout of the corridor leading to the interview rooms, and the use of the interview rooms themselves represent the power offender managers hold over their clients and the illusory aspect of probation being a sentence by consent. Finally, I look to the external aspect of the office and focus on the process of office consolidation which is entwined with both the rise of managerialsim as argued by Bottoms *and* the reformulation of probation as a punishment.

ARCHITECTURE AND CRIMINOLOGY

In the field of criminology, architecture has been used in various contexts. Looking at the architecture of a prison is perhaps the most obvious use of architecture in the criminological literature: after all, prisons have a physical structure for all to see. Some prison buildings hold more symbolic power than others: the gates of Wandsworth prison are recognisable to many British people, the roofs of HMP Strangeways are reminiscent of the days of prison riots, and the gates of St Alban's prison gatehouse bring back memories of the TV series *Porridge*. Similarly, the outlines of prison hulks sitting just off-shore remind people both of Charles Dickens' *Great Expectations* and the prisons crisis which occurred in the 1980s forcing the government to use ships to house prisoners off the coast of Britain. How prison buildings affect the behaviour within the prison has also been looked at with Sparks, Bottoms and Hay concluding that 'the design of prison buildings play a central part of in shaping the nature of the day-to-day routines' because the 'prison building is the instrument of incarceration'.[8] Jewkes and Johnston provide a detailed analysis of the evolution of prison architecture outlining the behavioural and symbolic importance of prison architecture:

> *the design of a prison impacts upon the lives of its occupants- inmates and staff- in a myriad of obvious and subtle ways. Furthermore… prisons stand as symbolic or allegorical statements of penal philosophy. Political judgements, policy priorities and public sentiments all play a role in the design, construction and location of penal institutions, and the symbolic and ideological forms with which prison buildings are invested have a vital role to play in explaining the internal power relations of the regime.*[9]

There have been several books specifically on architecture and the prison. Wortley focuses on how the use of the building, in situational crime control fashion, can be used to decrease the amount of crime which occurs *within* a prison.[10] Fairweather and McConville's edited volume focuses more on a description of prison architecture.[11] In reviewing Fairweather and McConville, Bosworth makes the point: '*that criminologists need to spend more time examining architecture, given that the design of a building will greatly affect those held within it*'.[12] It is clear, then, that the architecture of a prison has an accepted on the way a prison operates, both in terms of prisoner and staff experiences and the role the prison plays in the public's psyche.

Architecture is also an important aspect of situational crime prevention (SCP) which can be defined as '*a set of recipes for steering and channelling behaviour in ways that reduce the occurrence of criminal events*'.[13] SCP is more about the prevention of crime than the treatment of criminals. However, the use of architecture, in the form of CCTV cameras, higher and better fences around property, alarms, stronger reinforcements on entry points etc., is crucial

ISBN 978-0-9566139-1-2

to this practice and bears similarities to the architecture of prison described above. The inclusion of CCTV in the armoury of SCP suggests that architecture does not have to be restricted to 'bricks and mortar': that architecture can be more 'virtual'. Moreover, Jones[14] has used Lessig's concept of architecture[15] (or, as Lessig terms it, 'code') to look at Bottoms' model of compliance, SCP and electronic monitoring. Lessig's model allows us to see how physical constraints have an impact on behaviour in the context of criminology. More importantly for this paper, the model allows us to see how architecture '*can be seen as having values embedded within it, and how the social and communicative activities this code permits are thus influenced by the code's embedded values*'.[16] The way a building (or anything, for that matter) is designed holds some intrinsic value, which is then imparted onto the users of the building: '*code codifies values*'.[17]

ARCHITECTURE AND EDUCATION

Another area of academia that has looked at the role of architecture is education. Several pieces of research serve to highlight the importance of the physical environment of a school in terms of learning objectives and, importantly for this paper, discipline and compliance. In 1979 Weinstein conducted a review of the research on the impact of the physical environment of the school. The review found that the relationship between the physical environment and the school's pupils is present but complex. For example, students' choice regarding where they sit in a classroom can tell us both about their motivation to learn as well as their relationship with their peers; a relatively cluttered classroom can impact on the way students learn; and an 'ugly' classroom produced feelings of discontent, fatigue and a desire to escape amongst students.[18] Discipline connects the world of education with that of probation; and the role of the school building on the discipline which ensues is something which has long been recognised in the field of education. Dudek explains how the Victorian school house were designed to impose strict discipline on its pupils and how '*if the school is essentially concerned with imposing discipline, … this clearly neutralizes the potential for the school to develop creativity and expression as part of the educative process*'.[19] Although the purpose of a school is no longer to impose strict discipline, schools are still designed with discipline in mind and Dudek describes, for example, a school corridor in which the classroom entrances '*loop back on themselves to deter children from running into and out of classes at full speed*'.[20]

ARCHITECTURE AND PROBATION

It is clear from the discussion so far that architecture is more than simply how a building is designed but is about how a building or virtual constraint impacts on people's behaviour. I have also discussed how the way something is designed has a 'value' embedded within it that can tell us something about the era in which the design took place. Following on from this, probation can be seen to have an architecture in two ways. Firstly, the architecture of probation is evident in the very existence of probation offices and, secondly, probation can be perceived to have a more ephemeral type of architecture. Jones has already considered the second of these conceptualisations of probation architecture, as described above, and so the remainder of this paper considers the more physical aspect of probation's architecture. However, I look at this from a similar perspective to Jones and Lessig: that the way something is designed has an intrinsic value and an impact on the way people behave. There are similarities with Giddens' theory of structuration here,[21] although Lessig's conceptualisation relies on 'something' more tangible than the constraints of social structures in Giddens' theory. For the purposes of this paper, that 'something' is the design, layout and location of a probation building.

METHODOLOGY

This research consisted of 41 days of observations at a large probation office in the North of England, concentrating on how case managers interact with probationers, colleagues and policy as well as 'office life' in general. As part of the fieldwork I focused on compliance, discretion, managerialism, rehabilitation and risk management. In addition to, and as part of, this I observed how the architecture of the building creates, affects and perpetuates particular ways of working with probationers. This research, therefore, is a case study, the reliability, replicability and validity of which '*depends in large part on how far the researcher feels that these are appropriate for the evaluation of case study research*'.[22] As such, what I describe is not likely to, nor is it intended to, reflect the situation across the country. However, some of the features of 'my' probation office are typical of probation

ISBN 978-0-9566139-1-2

offices across the country: the CCTV, the locked doors, the protected reception space and the private interview rooms with panic buttons. What follows can therefore be described as *ideographic* with *nomothetic* potential: i.e. this analysis could have normative application when looking at probation work in different buildings and is important should new probation offices be designed. The case study described in this article is best seen as an 'exemplifying case' in that it provides a '*suitable context for certain research questions to be answered*'.[23] These research questions are: how does the architecture of a building affect the attitude of probation officers towards their clients? How might the architecture of a building affect the way probationers use and perceive a probation office? And, how does the architecture of a building reflect the rise of risk and punishment in probation policy?

THE PROBATION OFFICE

The research was conducted in an office which covers, in terms of geographical remit, one half of a large city in the north of England. The office houses six Offender Management Units (OMU) which comprise of, in total, between 55 and 60 Case Managers (CM), and Case Administrators (CA).[24] Each OMU is headed by a Senior Probation Officer (SPO) and there is one Practice Manager and an Assistant Chief Officer (ACO). It is worth mentioning that this office underwent a complete refurbishment several years ago and was therefore designed specifically for the purpose of probation in the early twenty-first century. The OMUs are arranged across two floors with the specialist services of Unpaid Work and Victim Services based downstairs and 'generic' OMUs upstairs. This OMU/specialist services split reflects the move away from specialist probation officers towards generic caseloads held by CMs after the implementation of the Offender Management Act 2007 in which specialist services, in the form of Accredited Programmes, were taken on by specialist programme tutors with CMs adopting on the role of case management rather than intervention providers.[25] The inclusion of a victim services unit reflects the 'remarkable return of the victim to centre stage in criminal justice policy'.[26] Despite this, 'the victim' was remarkably absent in my observations of CMs: there seemed to be an assumption that victims' needs would be dealt with by others. This physical split, could be seen as hindering the absorption of the victim into probation practice; although, arguably, policies around disclosure which prohibit CMs from contacting victims may be a greater barrier.

Although there are distinct teams in the form of OMUs, the office areas are open plan and thus it is difficult to discern who is in what team. The open plan nature of the office makes communication between members of staff easy and efficient and this collaboration was a distinct feature of the work. This collaboration was undoubtedly aided by the open plan design of the office and contrasts with the idea of the probation officer as 'technocrat' who must simply follow the rules and tick the boxes.[27] However, the design of the office also created tensions: for example, CMs are increasingly being encouraged to do programme reviews with probationers in prison by use of a telephone conference call and the noise in the office would sometimes make this difficult. The increased use of technology in the Probation Service which stems from resourcing issues (it is costly in terms of time and money to do a prison visit) is therefore being affected by the use of this open plan, collaborative design.

The SPOs' offices are located around the edge of the main office areas and have their own doors which are only closed during staff supervision sessions and when an SPO is accessing ViSOR.[28] Some SPOs share an office with another SPO. SPOs' offices tend to be located close to the team they oversee which makes communication between CMs, CAs and SPOs, again, easy and efficient. The privacy afforded by a separate office means that confidentiality can be guaranteed when necessary and that managers can do work without being disturbed if necessary. However, this separation also creates and perpetuates a feeling of us and them between management and frontline staff which might not be considered healthy in an organisation where managerialism can be described as hegemonic and the attitude towards management is, at times, negative. This separation was evident at times of team bonding when, for example, a birthday present was presented to a team member and the SPO was not invited, partly because no one could 'be bothered' to get him. Had a more inclusive approach to the placement of management been taken there would have been no choice but to include him in this activity.

All office areas are behind fob controlled doors so that no probationer could (or should) be ever in these areas. This means that staff are able to speak confidentially about clients. Moreover, as the building is surrounded by a fence, there is little possibility for paperwork to be seen from the street – something which became important during the period of research after NAPO reported that Phil Wheatley, the Director General of NOMS, had done just this.[29] The office areas represent, therefore, 'a zone of total confidentiality' which is clearly

ISBN 978-0-9566139-1-2

advantageous when dealing with sensitive data relating to offenders, victims and witnesses. However, this heightened level of confidentiality also allows staff to speak, generally in a jocular manner, derogatorily of their clients, be dismissive of their excuses for not attending and sceptical of their reasons for offending. Furthermore, the open plan aspect of the office allows people to spot oncoming management who may be critical of this way of speaking about probationers. This way of talking about an institution's clientele in confidential areas is not only seen in the context of probation and can be seen as either benign, as in the case of Crawley's[30] observation amongst prison staff and Waddington's[31] research in the police that 'humour is palliative';[32] or malign, as in the MacPherson report which states that 'the police canteen can too easily be [racism's] breeding ground'.[33] One CM suggested to me that this aspect of probation culture errs on the side of benign: 'it's about stress relief, I don't see it as a problem' (Fieldnotes). Nevertheless, the architecture of the building, and the fact that staff are so secure and separate from their clients enables and perpetuates this way of working.

Because of the nature of the office areas, it is easy to perceive of them as the 'backstage' of probation work because, 'the back region will be the place where the performer can reliably expect that no member of the audience will intrude'.[34] Goffman also argues that behaviour in the backstage is often inconsistent with what goes on frontstage,[35] and this is also reflected in the probation office. If the office areas are the backstage of probation then the waiting room and interview rooms are the frontstage which comprises of:

> *furniture, décor, physical layout, and other background items which supply the scenery and stage props for the spate of human action played out before, within, or upon it. A setting stays put, geographically, speaking so that those who would use a particular setting as part of their performance cannot begin their act until they have brought themselves to the appropriate place.*[36]

To express this in Goffman's terms is important because the frontstage '*functions in a general and fixed fashion to define the situation for those who observe the performance*'.[37] When discussing the problem of speaking derogatorily about clients arising from 'zones of total confidentiality' it is necessary to consider the '*relationship between culture and action*' and accept that '*if occupational culture is to serve as an empirically satisfactory concept as well as a theoretically necessary one, the sense of its internal variations and textures must be brought out*'.[38] The differences seen between behaviour in the front- and backstages is a way of achieving an empirically sound concept of probation culture. What happens in this arena defines, or at least contributes to the definition of, probation for the probationer themselves. When in the frontstage zone, staff are friendly and polite to offenders (although the nature of the job requires people to be firm and challenging at times). I would often see CMs greet people in the waiting room even if they weren't on their caseload and there was an air of mutual respect in terms of holding the main door open for others, for example. This idea of there being two sides to probation was backed up several times during the observation period with one CM saying, for example, that '*we are professional with the clients but not in the office*' (PO Female).

The waiting room is reasonably large and has about fifteen metal chairs. There are three toilets, a baby changing room and a water machine. The walls are painted a plain colour and are adorned with posters detailing the Service's responsibilities to offenders, motivational posters encouraging people to change, and advertisements of the different services on offer. That the waiting room is not completely sterile and functional suggests that probationers are not seen merely as members of a group that needs managing but are individuals who have their own basic needs beyond the need to reduce their offending. That probationers are seen as acceptable and, presumably fruitful, recipients of advertising also suggests a move towards seeing the probationer as customer and service user rather than offender. This could reflect Bottoms' theory that the increasing visibility of Unpaid Work might be justified on the basis of a 'new market state' although in this circumstance it is offenders instead of the community who have some element of 'consumer choice'.[39] The posters and generally amenable state of the waiting room might also be reflective of Phil Wheatley's recent use of the term 'Service User' in relation to prisoners and probationers.[40]

The interview rooms are located along a long, windowless corridor with security doors at several points along it (the initial door leading to the corridor is also controlled by a fob). It is not possible to move away from the waiting room without a fob but the reverse is possible. All the interview rooms except one have windows, often with the blinds drawn. They each have two doors – one leading on to the corridor with the other being an escape door leading to the next interview room. There are two panic buttons in each room – one by the main door and one on the desk. They all have a desk, a phone, a list of useful numbers on the wall and chairs (the number of which depends on the size of the room). Probationers are normally directed to sit at the end of the desk allowing

ISBN 978-0-9566139-1-2

the CM easy access to either door – there doesn't seem to be much consideration of the possibility that a probationer might want to leave in a rush. This pattern of directing the probationer where to sit can be interpreted as the CM exerting their control on a client in much the same way as Whyte describes a waitress or teacher's actions, '*so there is never any question as to who is in charge*'.[41] The setting, therefore, helps a CM to '*get off on the right foot*', and sets the tone for future interactions.[42] Goffman argues that '*the initial definition of the situation projected by an individual… has a distinctive moral character*'.[43] That one aspect of the initial definition is an imposition of control could imply that, despite many participants wanting to treat offenders with respect and equality (these are two of the 'official' values of this Probation Trust), this is not wholly possible within the physical structures of the office. The design of the interview rooms can therefore be seen as physical manifestations of this power differential between staff and offenders.

All interview rooms have CCTV which is viewable from reception (discussed below). CCTV in general doesn't feature much in the everyday life of probation work (although I observed one session in which the CM asked reception to keep an eye on the CCTV for the duration of the contact). However, once probationers enter the waiting room, corridor and interview rooms they are (potentially) under constant surveillance and the building design implies that they are not to be trusted. This reflects two aspects of probation practice. The first is that, in contrast to the waiting room, they are now treated as part of an aggregate group of offenders as described by Feeley and Simon in '*The New Penology*'.[44] The second is more metaphorical and reflects the historical idea of probation as a sentence by consent as well as the power held by a case manager. Although the requirement for offenders to consent to probation was repealed in the 1990s, the idea of consent is still important. NOMS have recently introduced the 'community compact' which offenders can sign at the beginning of an order. The compact lays out what each party can expect from the period of probation supervision but importantly the offender can refuse to sign. However, if they subsequently fail to comply with the Order they will be subject to breach proceedings as if they had consented. Probation staff also give offenders an illusion of consent and I observed this on several occasions – this was illusory because if an offender took up the offer of not complying they would face consequences in terms of poorly written pre-sentence reports or breach proceedings. Probationers are able to leave the building unhindered, reflecting the idea that they can, if they want, cease complying with the Service. However, when they leave the building they do not fall without the remit of the Probation Service and this alludes to the virtual side of probation's architecture: that of the 'invisible' forms of control that probation has over its clients such as home detention curfews, residency orders and warrants issued by the court for non-compliance. The physical layout of the corridor has important metaphorical allusions: to move forward without assistance is impossible whereas to leave without permission is feasible but ultimately illusory. A probationer may want to move forward in their lives but this can be slowed down by a variety of factors: a lack of capacity on programmes, a lack of flexibility in programme eligibility, or a lack of time on the part of the CM to spend sufficient time with someone to assess the risk that they pose and, if necessary to record any change.

The reception area is at one end of the waiting room with a large glass window separating the two. There are no holes in the glass: communication is done via a microphone system which the receptionist controls by pressing a button. When the button is not pressed the glass is surprisingly soundproof. There is a drawer at the bottom of the window through which documents can be passed, similar to those used in a bank. People in reception can therefore be seen by probationers, but not necessarily heard. Confidential conversations can occur in reception but there is a constant need to remember that a conversation could be overheard: it does not represent a zone of total confidentiality. Reception, therefore, is vulnerable to a 'backstage difficulty' which can result in someone knowing more about someone else than they perhaps should.[45] The backstage difficulty requires probation officers to adopt a frontstage performance, as seen in the waiting room, while simultaneously adopting a backstage persona. The design of the building, therefore not only disciplines the body of a probationer but also that of the case manager. This area of the probation building also represents the power held by probation over its offenders because they can be let in when staff want, and excluded when this is not necessary. That probation staff hold some element of power over their clients is not new and debates have occurred around the potential conflict of 'care and control' for many years. The staff I spoke to about this recognised the potential for this conflict but were simultaneously untroubled by it: they were happy to help their clients when they needed it but were also happy to '*shop 'em*' (Fieldnotes) when necessary. As well as holding some disciplinary potential over staff, the reception area can also be seen as a physical manifestation of this power differential between CMs and their clients.

ISBN 978-0-9566139-1-2

Having described the bricks and mortar of the probation building – how the layout inside the building allows for a split ontology when working with probationers, how CCTV allows for constant surveillance, doors create a sense of 'us and them', and how the tool of communication at reception gives staff the power to let a probationer in, or leave them out – we also need to consider the location of a probation building in order to examine how it is implicit in its aims and purpose. This is particularly important because we are in an era of probation work where home visits are increasingly rare,[46] offices in some areas (such as this one) have been consolidated to create large offices with correspondingly large geographical remits,[47] patchwork is non-existent, stringent National Standards mean that every probationer, at some point in their Order, will have to visit the probation office at least once a week for a period of time, and increasingly long sentences have resulted in people being on ever-longer licenses.

Bottoms has written about the consolidation of probation offices and argues that, ironically, while probation is more visible through initiatives such as Unpaid Work it is withdrawing from the community that it serves through the move to a '*smaller number of larger offices*'.[48] As more probationers are allocated to one office it allows for more efficiency and economies of scale so that savings can be made in terms of staff and building resources. This, as highlighted by Bottoms, is related to the rise of *'"new public management", including targets, key performance indicators, value for money etc.'*,[49] and so can be seen to be reflective of the hegemony of managerialism in NOMS.[50] Having considered that group interventions have been proven successful in reducing reoffending,[51] Bottoms comes to the conclusion that '*if a consequence of that investment [in group programmes] is a more office-based probation service, so be it*'.[52] However, Bottoms also argues that this consolidation means that officers rarely visit the areas in which offenders live and so '*office-based staff… are unlikely to inspire such confidence*' in offenders of their ability to assist them in dealing with the issues they face.[53] My observations confirm that not only do officers rarely visit the communities they serve, but also that there is an element of fear in visiting them: on one occasion a PSO took her name badge off before going into a supermarket, saying that they didn't want to risk drawing attention to themselves; and another PO described how when he walked to an offender's house (who he was visiting purely because the offender had broken his ankle) he would not take any valuables with him, would memorise the route so as not to have to look at a map and look confident at all times. The CMs in my office were clearly not well known in the community and were happy to stay anonymous – a far cry from the 1960s when the probation officer would be well known on their 'patch'.[54] The programme of consolidation has clearly resulted in probation becoming more removed from the communities it serves.

The office is on a main road, served by numerous bus routes, in a densely populated, deprived area of the city. Clearly, many clients live nearby and others live a short bus ride away. However, others have to get more than one bus and travel for almost an hour. Some walk because they don't have enough money to pay for the bus in the first place. Despite punishment not being an explicit aim of probation in the 1960s it was recognised that, on the part of the probationer, there is '*a punitive element in the requirement to submit to supervision*'.[55] The consolidation programme will undoubtedly have increased the distance some probationers will have to travel and this adds an additional burden to an offender depending on where they live. Durnescu has recently described how one of the '*pains of probation*' is the distance probationers have to travel to their probation appointments; this was considered 'more painful' by those who had to expend more time and money in order to successfully comply with their probation order.[56] The situation of a probation building therefore affects the 'amount' of punishment a probationer receives, depending on their place of residence. The process of consolidation increases the level of punishment inflicted to certain people depending on where they live. The programme of office consolidation therefore reflects not only the rise of the 'new public management' but also the reformulation of 'probation as punishment in the community' in the 1990s.[57]

CONCLUSION

Through reference to my experiences at a probation office in England I have shown that the building both reflects recent changes in probation policy and purposes and that the building plays an integral part in how staff treat offenders. By drawing on Goffman, probation staff create and present a definition of probation to their clients which is inextricably linked to the 'setting' of a probation office. We have seen that the backstage of probation work allows officers to speak freely about their clients but that this creates a frontstage and backstage of probation work. We have also seen how technologies such as telephone conferencing are affected by the design of a building and how technologies such as CCTV reflect the increased role of risk in the work of

ISBN 978-0-9566139-1-2

probation officers. Finally, I have shown how the practice of office consolidation both separates staff from the communities they are supposed to be serving and increases the burden of probation resulting in an increased level of punishment for some.

I suggest, therefore, that by looking at how the architecture of probation affects the people who use it, we can learn more about what probation is and what probation does. This kind of work, embedding practice in its physical surroundings, might also be useful if the policy of consolidation is reversed (as was suggested to me by one SPO at the office) and local, smaller probation offices are reopened. If we take Gill's proposition in relation to prisons that the building must be the means by which an institution's aims are met,[58] it becomes clear that if we are witnessing a change in probation policy towards more locally delivered probation work we need to take a fresh look at probation offices with a view to making them more conducive to achieving the different aims of probation. If, for example, the importance of compliance over enforcement continues and policy makers come to the conclusion that increased compliance relies on the relationship between probationer and case manager (as suggested by McNeill and Robinson),[59] then to create a building which creates barriers and tension between the two will potentially undermine any positive impact of such a change in policy.

REFERENCES

[1] Gelsthorpe, L., & Morris, A. (2002). Women's imprisonment in England and Wales: A penal paradox. *Criminology and Criminal Justice*, *2*(3), 277-301.
[2] Goffman, E. (1971). *The Presentation of Self in Everyday Life*. Pelican books. Harmondsworth: Penguin.
[3] Bottoms, A. E. (2008). The Community Dimension of Community Penalties. *The Howard Journal of Criminal Justice*, *47*(2), 146-169.
[4] There are 85,000 people in prison and 250,000 people under the supervision of the Probation Service at any one time (Ministry of Justice, C. D. (2009). *Offender management caseload statistics 2008*. Statistics Bulletin. London: Ministry of Justice. Retrieved from http://www.justice.gov.uk/publications/docs/offender-management-caseload-statistics-2008-2.pdf)
[5] Gelsthorpe, L., & Morris, A. (2002). Women's imprisonment in England and Wales: A penal paradox. *Criminology and Criminal Justice*, *2*(3), 277-301. p.292.
[6] Armstrong, S. C., & McAra, L. (2006). Audience, borders and architecture: the contours of control. In L. McAra & S. C. Armstrong (Eds.), *Perspectives on Punishment: The Contours of Control*. Oxford: Oxford University Press. pp. 23-24.
[7] Robinson, G. (2008). Late-modern rehabilitation: The evolution of a penal strategy. *Punishment & Society*, *10*(4).
[8] Sparks, R., Bottoms, A. E., & Hay, W. (1996). *Prisons and the problem of order*. Oxford University Press., p. 229.
[9] Ibid., 191.
[10] Wortley, R. (2002). *Situational prison control: Crime prevention in correctional institutions*. Cambridge University Press.
[11] Fairweather, L., & McConville, S. (2000). *Prison architecture: policy, design, and experience*. Oxford: Architectural Press.
[12] Mary Bosworth, "Prison Architecture: Policy, Design and Experience. Edited by Leslie Fairweather and Sean McConville (Oxford: Architectural Press, 2002, 162pp. {pound}50.00 hb)," *Br J Criminol* 43, no. 3 (2003): 635.
[13] Garland, D. (2000). Ideas, institutions and situational crime prevention. In A. von Hirsch, D. Garland, & A. Wakefield (Eds.), *Ethical and Social Perspectives on Situational Crime Prevention* (pp. 1-16). Oxford: Hart Publishing. p.1.
[14] Jones, R. (2006) Architecture, Criminal Justice, and Control. In L. McAra & S. C. Armstrong (Eds.), *Perspectives on Punishment: The Contours of Control*. Oxford: Oxford University Press.
[15] Lessig, L. (1999). *Code and Other Laws of Cyberspace*. New York, NY: Basic Books.
[16] Jones, R. (2006) Architecture, Criminal Justice, and Control. In L. McAra & S. C. Armstrong (Eds.), *Perspectives on Punishment: The Contours of Control*. Oxford: Oxford University Press. p.178.
[17] Lessig, L. (1999). *Code and Other Laws of Cyberspace*. New York, NY: Basic Books. p.59.
[18] Weinstein, C. S. (1979). The Physical Environment of the School: A Review of the Research. *Review of Educational Research*, *49*(4), 577-610.
[19] Dudek, M. (2000). *Architecture of Schools: The New Learning Environments*. Oxford: Architectural Press. p.10.
[20] Ibid. p.55.
[21] *Giddens, A. (1984). The constitution of society : outline of the theory of structuration. Berkeley: University of California Press.*
[22] Bryman, A. (2004). *Social Research Methods* (2nd ed.). Oxford: Oxford University Press, p. 50.
[23] Ibid., 51.
[24] I use the term Case Manager to refer to what the Offender Management Act terms the Offender Manager. This includes both Probation Officers and Probation Service Officers. A Case Administrator is an unqualified person who deals with sending letters out, creating reports on the computer system, collating statistical data, allocating appointments and other administrative duties.

[25] Raynor, P., & Vanstone, M. (2007). Towards a correctional service. In L. Gelsthorpe & R. Morgan (Eds.), *Handbook of Probation*, p. 74.
[26] Garland, D. (2001). *The Culture of Control: Crime and Social Order in Contemporary Society*. Oxford: Clarendon, p. 11.
[27] Bailey, R., Knight, C., & Williams, B. (2007). The Probation Service as Part of NOMS in England and Wales: Fit for Purpose? In R. Morgan & L. Gelsthorpe (Eds.), *Handbook of Probation* (pp. 114–30). Cullompton: Willan Publishing.
[28] ViSOR is a secure system that has been developed to support Multi-Agency Public Protection Arrangements (MAPPA). It can be accessed by the Police, Probation and Prison services, enabling the sharing of risk assessment and risk management information on individual violent and sex offenders (http://noms.justice.gov.uk/news-publications-events/publications/impact-assessments/ViSOR_EIA_08).
[29] Fletcher, H. (2009, November). Wee Philly Wheatley. *NAPO News*, *214*, p. 3.
[30] Crawley, E. M. (2004). Emotion and performance: Prison officers and the presentation of self in prisons. *Punishment Society*, *6*(4).
[31] Waddington, P. A. J. (1999). Police (canteen) sub-culture. An appreciation. *British Journal of Criminology*, *39*(2).
[32] Crawley, E. M. (2004). Emotion and performance: Prison officers and the presentation of self in prisons. *Punishment Society*, *6*(4), p. 419.
[33] The Stephen Lawrence Inquiry. Retrieved June 2, 2010, from http://www.archive.official-documents.co.uk/document/cm42/4262/4262.htm. Para. 6.17.
[34] Goffman, E. (1971). *The Presentation of Self in Everyday Life*. Pelican books. Harmondsworth: Penguin, p. 116.
[35] Ibid., p. 118.
[36] Ibid., p. 32-33.
[37] Ibid., p. 32.
[38] Fielding, N. (1988). Joining forces: police training, socialization, and occupational competence. London: Routledge, p. 185.
[39] Bottoms, A. E. (2008). The Community Dimension of Community Penalties. *The Howard Journal of Criminal Justice*, *47*(2), 146-169. p. 155.
[40] James, E. (2010, February 24). Striking a balance between punishment and compassion. *The Guardian*. London. Retrieved from http://www.guardian.co.uk/society/2010/feb/24/phil-wheatley-national-offender-management-service
[41] Whyte, W. F. (1946). When workers and customers meet. *Industry and Society*, pp.132-133 cited in Goffman, E. (1971). *The Presentation of Self in Everyday Life*. Pelican books. Harmondsworth: Penguin, p. 23.
[42] Goffman, E. (1971). *The Presentation of Self in Everyday Life*. Pelican books. Harmondsworth: Penguin. p. 23.
[43] Ibid., p. 24.
[44] Feeley, M. M., & Simon, J. (1992). The New Penology: Notes on the Emerging Strategy of Corrections and Its Implications. *Criminology*, *30*(4), 449-474.
[45] Goffman, E. (1971). *The Presentation of Self in Everyday Life*. Pelican books. Harmondsworth: Penguin. p.121-122.
[46] Home visits are only mandatory for Tier 2, 3 and 4 probationers who pose a high or very high risk of harm and, even then, only once in the duration of the Order or license period. Beyond the requirements laid out in the National Standard, the participants in my study appeared to use home visits primarily to increase compliance, rather than the more traditional usage to get a broader picture of a client's life.
[47] Bottoms, A. E. (2008). The Community Dimension of Community Penalties. *The Howard Journal of Criminal Justice*, *47*(2).
[48] Ibid., p. 161.
[49] Ibid., p. 162.
[50] Bailey, R., Knight, C., Williams, B., Morgan, R., & Gelsthorpe, L. (2007). The Probation Service as Part of NOMS in England and Wales: Fit for Purpose? In *Handbook of Probation* (pp. 114–30). Cullompton: Willan Publishing, p. 115.
[51] P. Raynor, "Rehabilitative and reintegrative approaches," in *Alternatives to Prison: Options for an Insecure Society*, ed. A. Bottoms, S. Rex, and G. Robinson (Cullompton: Willan Publishing, 2004), 195–223.
[52] Bottoms, A. E. (2008). The Community Dimension of Community Penalties. *The Howard Journal of Criminal Justice*, *47*(2).
[53] Ibid.
[54] cf. St John, J. (1961). *Probation, the Second Chance*. Vista Books. & Jarvis, F. V. (1972). *Advise, Assist and Befriend: A History of the Probation and After-Care Service*. London: National Association of Probation Officers.
[55] Morison Report. (1962). *Report of the Departmental Committee on the Probation Service*. London: H. M. Stationery Office, para. 11.
[56] Durnescu, I. (2010). Pains of Probation: Effective Practice and Human Rights. *International Journal of Offender Therapy and Comparative Criminology*. vol. 0.
[57] Raynor, P., & Vanstone, M. (2007). Towards a correctional service. In L. Gelsthorpe & R. Morgan (Eds.), *Handbook of Probation*, p. 67.
[58] Gill, H. B. (1962). Correctional Philosophy and Architecture. *Journal of Criminal Law, Criminology and Police Science*, *53*, 312.
[59] Burnett, R., & McNeill, F. (2005). The place of the officer-offender relationship in assisting offenders to desist from crime. *Probation Journal*, *52*(3), 221. Robinson, G., & McNeill, F. (2008). Exploring the dynamics of compliance with community penalties. *Theoretical Criminology*, *12*(4), 431.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

MUSINGS ON MATHS AND MUSIC – THE COMPLEX PATTERNS WE SIMPLY DON'T HEAR

Michelle PHILLIPS

PhD Student, University of Cambridge, Cambridge, UK
mep41@cam.ac.uk

Keywords: Music, Mathematics, Golden Section, Musical Time

ABSTRACT

We're all familiar with the popular belief that maths and music are related, perhaps through anecdotal childhood experience, perceived competency in friends or colleagues, or familiarity with modern compositional techniques. And indeed a combination of the two disciplines may proffer various strands of interest – the mathematics of sound vibration, the logic of the Western system of musical notation, or J. S. Bach's embedded numerical codes.

This paper will focus on the field of music analysis, where it is often claimed that mathematical relations exist in the musical score. For example, there are more than 130 separate examples of music analysis which incorporate the golden section or Fibonacci sequence, a popular concept since the 1970s owing largely to the work of music analysts Lendvai and Howat.

Though such an exploration mathematical structuring used in musical composition may be informative about the composer's own musical conceptions, aural perception of a musical work as it unfolds in time is governed by vastly more intricate principles. Compared to the work as experienced during music analysis, the listener does not perceive the musical bars in objective and absolute relation to one another. Rather, as extensive studies of psychological time have demonstrated, experience of duration depends on multiple complex factors, such as attention, memory, context of the event, level of arousal, preference and familiarity. Attention to the nature of the experience of musical time is relatively new and as yet under-explored; however, it is vital to any consideration of a potentially aurally detectable mathematical relation.

I will argue that, whilst attribution of a musical work's aesthetic value or canonical significance to its underlying mathematical structure may be alluring, there are no grounds for such claims, as we simply don't hear such complex patterns.

INTRODUCTION

Finding structures in notation tells us very little about how minds experience the music.
– Professor David Huron (keynote speaker), ESCOM 2009 7th Triennial Conference of the European Society for the Cognitive Sciences of Music

My music teacher would sigh and smile under his faked frown as I turned up late again for A-level music classes after maths lessons. My maths teacher would do the same on the days when the lessons were reversed. And on the rare occasion that the two teachers coincided in my presence, they'd chat and revel in the apparent verification of their ongoing theory - that students who were good at maths or music were also good at the other.

This anecdote will, I'm sure, be familiar to many. The belief that music and maths somehow belong together is popular and widespread, and shows no sign of abating; books with various music-mathematical titles continue to

ISBN 978-0-9566139-1-2

grace our bookshops and library shelves, with *The Math Behind the Music*, *Music: A Mathematical Offering* and *Music and Mathematics: From Pythagoras to Fractals*[1] published in the UK in the last seven years alone.

Yet these books are not variations on the same theme. Their subject matter ranges from acoustics, to musical composition, to the history of the teaching of the two subjects from the age of the medieval quadrivium. The notion of music and maths, or maths and music, does not represent one strand of thinking, but is rather a broad term for a number of different areas of study.

Largely, the areas of study or interest, whether academic or popular, which tend to form part of what is generally conceived of as 'music and maths' can be grouped into the following six categories:

1. Educational abilities and development (the belief that people who are good at maths are often also good at music; the widely discredited 'Mozart effect')
2. Neuroscientific invetigation (the processing of music and maths in the brain)
3. Musical notation of pitch, rhythm and metre (the idea that music is purely made up of maths, or *is* maths)
4. The physics of sound (acoustics, tuning systems and the engineering of musical instruments)
5. Compositional techniques (modern composers' frequent use of mathematical patterns and concepts to structure compositions)
6. Music working *through* maths (the assertion that music owes its power and / or emotional effect to its underlying mathematics or patterns; the idea that you can hear mathematical patterns in music, and these contribute to the listener's enjoyment)

This paper will touch briefly on all of these categories, but will focus largely on the sixth. The mystery and allure of the pairing of music and maths has reached a new level in recent decades, with today's mathematicians claiming music works through maths, owing its effect, and / or affect, to its underlying mathematics. Such claims fail to take into account the large body of research which has emerged over the last 30 years regarding music perception. This growing field of study demonstrates that perception of music is governed by multiple, complex, and as yet under-explored psychological processes. Hence, regardless of whether a musical *score* appears to contain fascinating and intricate patterns (whether intentional on the composer's part or not), such relationships will not necessarily be communicated to the listener when the musical work is aurally realised *in time*. Any claim that music's effect, in part or in whole, owes its worth in any way to its mathematical structure must take into account how this music is perceived by the listener, and not rely solely on score study.

MUSIC AND MATHS

The history of music and maths is long and winding, from Pythagoras' assertion that harmonious note relationships originate in simple whole number ratios (e.g. dividing a string in the ratio 2:1 produces an octave, 2:3 a major fifth, and 3:4 a major fourth and so on), to composer Phil Thompson's 'theme tune of the Universe', which translates images of fractals into music[2], to Wikipedia's extensive entry on 'Music and Mathematics'. [3]

In 2007 a maths and music specialist school, 'The Bridge Academy', supported by the finance company UBS, opened in East London. This academy claims to be based on the idea that 'Fundamentally, music and mathematics complement each other'[4] and furthermore 'Educationalists will tell you the subjects are related'. Despite these bold claims and the significant financial investment they have brought about, there is currently no evidence that an aptitude in one of these two disciplines leads to aptitude in the other, and statements to the effect 'people who are good at music are also good at maths' (or vice versa), almost invariably begin 'I've heard' or 'it's rumoured', giving no evidence for the validity of such a claim.

Prominent mathematicians are also guilty of overplaying the evidence of a link between music and maths in education – one eminent scholar recently claimed on breakfast television that GCSE candidates would be advised not to listen to music whilst studying, unless they were learning maths, as the two subjects utilise the same parts of the brain. There is no evidence for this claim, and indeed fMRI studies designed to examine how music is processed neurologically have not found any one specific brain area alone to be involved in musical engagement – music appears to involve many neurological processes concurrently.

ISBN 978-0-9566139-1-2

The temptation to exaggerate the links between music and maths is hence great enough for beliefs within the area to have become popular and widespread, despite the lack of any empirical verification. Yes, there are number relationships at the very heart of the Western music tradition, with its system of crotchets, quavers, minims, metres from 4/4 to 3/4/ to 6/8, and pitches which divide the octave into 12 equal steps, but such numerical relationships are not necessarily transmitted to the listener. A distinction must be maintained between the conceptual relationships evident through music analysis, and their perceptual counterparts.

MUSIC PERCEPTION

Over the last 30 years, study in the field of music perception (usually under the banner of 'music and science' or 'music psychology') has blossomed. [5] Although the field is young, its empirical nature makes it ideally suited to potentially shed light on many popular beliefs relating to the perception of musical structure and form.

One core area of research within music perception is how musical form is perceived as it unfolds in time. Empirical work undertaken to date in this area has suggested that musical form may not in fact be perceived globally (i.e. perception of the entire musical work) at all, but rather only on the local level (i.e. moment to moment, for example a single phrase or motif), [6] and that altering a work's tonal structure may go unnoticed by trained and untrained musicians alike. [7] Other musical features may take priority in perception over global structure: for example, a study by Deliège et al.[8] finds weak sensitivity to large-scale form compared to cues perceived on the music's surface. However, studies have also indicated some perception of large-scale form by musicians and non-musicians, [9] finding that listeners demonstrated highly veridical judgments of location of a musical extract within the whole, and that there was high inter-subject agreement about the features marking segment boundaries. [10]

Hence the question of whether we perceive the entire piece of music at all is still very much under debate, rendering assertions of perception of large-scale mathematical structuring quite unjustified.

The second section of this paper will take one mathematical concept, which is widely held by music analysts to appear in large scale musical form, as an example of a mathematical structure purported to contribute to a work's aesthetic effect, but for which there is no evidence of aural perception. This mathematical concept will be used to illustrate how there is a very real tendency in the fields of both maths and music to overplay the relationship between the two disciplines – the golden section is a prime example of a complex pattern we simply don't hear.

MUSIC, MATHS AND THE GOLDEN SECTION

The golden section is obtained by dividing a whole into two parts such that the ratio of the smaller to the larger section is equal to that of the larger to the whole (BC:AB = AB:AC):

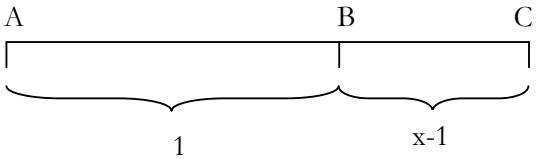

Figure 1: Dividing a whole in the ratio of the golden section.

Solving the resulting quadratic equation ($x^2 - x - 1 = 0$) gives a ratio of 1:1.618 to three decimal places, a ratio termed the 'golden section'.

Also known as the golden mean, ratio, number and divine proportion, and arithmetically linked to the Fibonacci series, the golden section possesses intriguing mathematical properties. Yet its popularity extends beyond mathematics: the golden section is represented in nature (for example, in seed formation in pine cones and sunflowers); the field of experimental psychology continues to critically examine and build on Gustav Fechner's work with 'golden rectangles'[11] and to address the question of whether the golden section is indeed an artefact of

ISBN 978-0-9566139-1-2

visual aesthetic preference; social psychologists claim human judgments of positive and negative attributes in others conform to this ratio; and dentists are taught the aesthetic appeal of teeth proportioned according to the golden section during training. Popular culture also devotes attention to the concept: the social networking website Facebook includes the groups 'oh beautiful 1.618, please don't ever terminate' and 'I Love The Golden Ratio and Fibonacci Numbers'; and the online micro-blogging service Twitter includes members using the identifiers 'GoldenRatio' '1x62 Golden Ratio', 'Fib_numbers' and 'Sezione_Aurea', amongst many others bearing relation to the golden section.

Over the last 30 years the golden section has received increasing attention in the field of music analysis, encompassing music from Gregorian chant[12] to Madonna.[13] This popularity is largely owing to Ernö Lendvai's[14] work on the music of Béla Bartók, first published in English in the 1970s, and Roy Howat's[15] subsequent monograph, *Debussy in Proportion.*

There are multiple examples of composers stating their intentional use after this and into the 21st century. Post-1900, and largely post-Lendvai and Howat's landmark studies, this technique was widely employed (composers include Hindemith, Xenakis, Schoenberg, Ligeti, Stockhausen and Nono), and continues to be used extensively today (for example by composers such as Murail, Gubaidulina, Tutschku, Knussen, Maxwell Davies, Holliger, Wuorinen, Sáry, Weir and Turnage). Nevertheless, claims by composers that they have successfully manipulated the listener's aural experience as the music unfolds in time remain unsubstantiated, though they may be informative about the composer's own musical conceptions.

There are more than 130 separate examples of music analysis which incorporate the golden section, all with one surprising uniting factor – not one of these studies tackles the golden section in music empirically. That is, although these studies claim to identify the proportion within the musical notation, and often make generic claims regarding the aesthetic value lent to the work by the manifestation of this ratio, not one demonstrates that this ratio may be heard. Indeed, few consider this potential perceptual obstacle to such conclusions. It is hence not known whether the golden section is aurally perceptible, nor has any study yet attempted to address this

Less than a quarter of these 126 documents of music analysis acknowledge the question of whether the golden section may be perceived during music listening, rather than being retrievable through score analysis. Hence claims such as '[Debussy's proportional systems] show ways in which the forms are used to project the music's dramatic and expressive qualities with maximum precision?'[16] which assert that the golden section lends aesthetic value to a composition when it is heard, are unfounded – if there is no evidence the golden section is indeed aurally salient, how can it's employment, whether conscious or not on the part of the composer, lend aesthetic worth?

The allure and mysticism surrounding the golden section has spurred many so-called 'golden ratio enthusiasts' or 'believers in the golden section'. Music analysis often demonstrates careless calculation of purported mathematical bases, or falls foul of confirmation bias as those guilty of heavy-handed 'Lendvain thinking' find what they have set out to find. In Donald Tovey's words 'there are so many ways of taking your sections that I doubt whether any musical composition can avoid golden ones somewhere'. [17]

So how can we begin to investigate whether the golden section, or in fact any purported mathematical relationship in a musical work, may be aurally detectable by the ear in a musical work's realisation?

MUSICAL TIME

W hy do we always have to think of music in terms of notes? ... We work with sounds, for which notes are simply symbols... Notes and sounds are not the same thing.

- Anderson, J., & Murail, T. "In Harmony. Julian Anderson Introduces the Music and Ideas of Tristan Murail". (1993) *The Musical Times 134* (1804), p. 321.

Unlike visual art, music is transitory, fleeting, and only available to human perception in the moment. It is

ISBN 978-0-9566139-1-2

experienced in what Fraisse may call the 'perceptual present',[18] and is subject to the limits of human working memory, thought to be between 5 and 8 seconds.[19] Hence appraisal of music's global structuring and proportions may only be undertaken in retrospect, once the piece has reached its conclusion. Again, such cognitive appraisal relies on memory resources, which are indeed limited, and which would be unlikely to retain an accurate linear record of an entire work.
The most prevalent form of golden section analysis identifies a significant point in the musical work and claims to have discovered, by counting bars, beats or notes, that this point divides the work in the ratio of the golden section. Over 90% of golden section analysis takes this form and employs this methodology. Hence, essentially, golden section analysts claim that one section of the work exists in relation to the other in the ratio of the golden section. To assume that this divide may be *perceived* is to assume that one section of music is perceived in absolute relation to another.

However, compared to how a work may be experienced during music analysis, the *listener* does not perceive the musical bars or sections in objective and absolute relation to one another. Rather, as extensive studies of psychological time have demonstrated,[20] daily experience of events in time depends on multiple complex factors, such as context of the event, level of arousal, body temperature, preference and method of timing estimation; in other words, a watched kettle does take longer to boil, time does fly when you're having fun, and the years do pass more quickly as you age.

'Time' may be thought of as the seconds, minutes, days and years by which our routines and lives are regulated. This is clock time, measured according to the radioactive decay of the caesium atom in the atomic clock. However, this objective form of measuring and controlling events in time is not representative of the form in which humans experience such events. Rather, psychological time is pliable and subjective, and the speed at which we experience events varies according to multiple factors, including context, level of arousal, preference, pharmacological influence, body temperature and engagement in the task at hand.

The question of whether the golden section can be perceived aurally, in addressing the listener's experience of length of musical sections in relation to one another, necessitates an understanding of the notion of musical time, that is the subjective experiences of elapsed duration during music listening. During the last 20 years, empirical work in *music* psychology has begun to explore these theories of psychological time, and has suggested that musical features (such as harmonic modulation, tempo variation, or tonality) may alter the listener's perception of elapsed duration of a section of music.[21]

THE EMPIRICAL STUDY OF MUSICAL TIME

Musical time is a new, and as yet relatively little understood field. It remains to be explored whether musical time conforms to popular models of psychological time, be they psychological (for example the internal clock or contextual change models) or neurological, in other words determined by neural oscillations, or whether, as Mari Riess Jones proposes, 'temporal production and estimation have special meanings in musical contexts'. [22]

My results relating to musical time are still in their early stages, but these and other emerging studies demonstrate that systematic manipulation of music's modularity, familiarity, or tempo may alter a listener's estimation of elapsed duration. For example, my manipulation of the speed of the familiar Nokia ringtone leads experimental participants to let the phone ring longer before answering when the motif is slowed by 40% of the original speed. Reducing the speed of a familiar ringtone makes people feel less time has elapsed. Hence, if, as is often the case, a musical work or section were to contain different tempi, modalities, or levels of familiarity to the listener in one part, in comparison to the other, the elapsed duration of each part will likely be experienced in different ways. These durations will relate to each other only in subjective terms, and in no circumstance will one musical bar in section A of that work be experienced as absolutely relative, in terms of duration, to section B.

Thus experience of time is subjective, and in its reliance on context, association, physical and chemical condition, and emotional arousal, amongst other factors, experience of elapsed duration is different for every individual, and for every experience. Hence claims that the golden section 'contributes to our sense of aesthetic satisfaction',[23] or indeed that such complex mathematical patterns, which may be evident through score analysis, are aurally identifiable or responsible for the aesthetic or affective power of the music, are entirely unfounded.

ISBN 978-0-9566139-1-2

During music listening, such complex relationships are subject to the limits of psychological time, and are thus simply not aurally perceived.

CONCLUSION

My enthusiasm about the possible connections between these two disciplines has, I hope, been demonstrated in the above musings, but only in conjunction with my skepticism. I hope to have demonstrated, using the example of the golden section, that complex mathematical patterns which may be evident in a musical score, or intentional on the part of the composer, are simply not audible to the listener as the music unfolds in time.

Whilst mathematical investigation of a musical work may be informative about the composer's own musical conceptions, there is no justification for the assertion that music works *through* maths, or that rigorous mathematical structuring lends aesthetic or affective value to the musical work. As exploration of the notion of psychological time demonstrates, we simply don't hear such complex patterns.

REFERENCES

[1] Harkleroad, L. *The Math Behind the Music.* (2006) Cambridge: Cambridge University Press, Benson, D. *Music: A Mathematical Offering.* (2007) Cambridge: Cambridge University Press, Fauvel, J., Flood, R., & Wilson, R. J. *Music and Mathematics: From Pythagoras to Fractals.* (2003) Oxford: Oxford University Press.
[2] www.fractalmusician.com
[3] Wikipedia, "Music and Mathematics", http://en.wikipedia.org/wiki/Mathematics_and_Music
[4] The Bridge Academy, http://www.bridgeacademy.hackney.sch.uk/FAQ.htm#anchor8
[5] The field is widely believed to have begun around the time of Lerdahl and Jackendoff's *A Generative Theory of Tonal Music*, published in 1983 (Cambridge, Mass; London: MIT Press), although there are those who would assert Leonard Meyer's 1956 *Emotion and Meaning in Music* (Chicago: Chicago University Press) was the true beginning.
[6] Tillmann, B. and Bigand, E. "The Relative Importance of Local and Global Structures in Music Perception". (2004) *The Journal of Aesthetics and Art Criticism 62* (2), p. 211-22.
[7] Cook, N. "The Perception of Large-Scale Tonal Closure". (1987) *Music Perception 5*, p 197-206.
[8] Deliège, I., Melen, M., Stammers, D., & Cross, I. "Musical schemata in real-time listening to a piece of music". (1996) *Music Perception 14* (2), p. 117-159.
[9] Lalitte, P., & Bigand, E. "Music in the moment? Revisiting the effect of large scale structures". (2006) *Perceptual and Motor Skills 103* (3), p. 811-828, Bigand, E. "More about the musical expertise of musically untrained listeners". *Paper presented at the Conference on Neurosciences and Music: Mutual Internactions and Implications on Developmental Functions, Venice, Italy* (2002, Oct 25-27).
[10] Clarke, E. F., and Krumhansl, C. L. "Perceiving Musical Time". (1990) *Music Perception 7* (3), p. 213-52.
[11] Fechner, G. T. "Various Attempts to Establish a Basic Form of Beauty: Experimental Aesthetics, Golden Section, and Square". (1997, translated from 1876) *Empirical Studies of the Arts 15* (2), p. 115-130.
[12] Larson, P. "The Golden Section in the Earliest Notated Western Music". (1978) *Fibonacci Quarterly 16* (6), p. 513-515.
[13] Cook, N. *Analysing Musical Multimedia.* (1998) Clarendon Press, Oxford, p. 166.
[14] Lendvai, E. *Béla Bartók: An Analysis of His Music.* (1971) Kahn and Averill, London.
[15] Howat, R. *Debussy in Proportion.* (1983) Cambridge University Press, Cambridge.
[16] Howat, R. *Debussy in Proportion.* (1983) Cambridge University Press, Cambridge, p. 1.
[17] Tovey, D. F. *Essays in Musical Analysis, Volume I-II, Symphonies.* (1935) London: Oxford University Press, p. 19.
[18] Fraisse, P. *The Psychology of Time*, Jennifer Leith, trans. (1963) New York: Harper & Row.
[19] Baddeley, A. D. *Working Memory, Thought, and Action.* (2007) Oxford: Oxford University Press.
[20] For detail see Block, R. A. "Models of Psychological Time". In R. A. Block (Ed.), *Cognitive Models of Psychological Time* (pp. 1-35). (1990) Hillsdale, New Jersey: Lawrence Erlbaum Associates.
[21] See for example: Firmino, É. A., Bueno, J. L. O. et al. "Travelling Through Pitch Space Speeds up Musical Time". (2009) *Music Perception 26* (3), p. 205-209., Kellaris, J. J., & Kent, R. J. "The Influence of Music on Consumers' Temporal Perceptions: Does Time Fly When You're Having Fun?" (1992) *Journal of Consumer Psychology 1* (4), p. 365-376, Noseworthy, T. J., & Finlay, K. "A Comparison of Ambient Casino Sound and Music: Effects on Dissociation and on Perceptions of Elapsed Time While Playing Slot Machines". (2009) *Journal of Gambling Studies 25*, p. 331-342.
[22] Jones, M. R. "Musical Events and Models of Musical Time". In R. A. Block (Ed.), *Cognitive Models of Psychological Time* (pp. 207-238). (1990) Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., p. 214.
[23] Atlas, A. W. "Stealing a Kiss at the Golden Section: Pacing and Proportion in the Act I Love Duet of la Bohème". (2003) *Acta Musicologica 75* (2), p. 269-291.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

LANGUAGE ARTS AND VISUAL ARTS: BRIDGING THE GAP WITH VISUAL LITERACY

Catherine READ

PhD Student, Simon Fraser University, British Columbia, Canada
read@sfu.ca

Keywords: Literacy, Reading, Writing, Visual art, Picture book, Education, Teaching

ABSTRACT

In the field of Education, Language Arts and Visual Arts are usually considered to be at opposite ends of the curriculum. An emphasis on Language Arts in schools is common because it focuses on the key skills of reading and writing. Language Arts maintains an aura of importance and is offered daily to students. However, Visual Arts are usually at the other end of the education pendulum. This subject is often optional, offered to students sporadically, and is usually considered to be an educational frill.

This paper examines the juncture between these two seemingly opposite disciplines. Can experience in the Visual Arts aid the acquisition of Language Arts skills? Participation in the Visual Arts, both as a viewer and a creator, can expand one's visual literacy skills. Visual literacy in turn, can aid in developing a secure foundation that is needed for reading and writing. These ideas are examined through an exploration of current literature and contemporary visual arts programs. Practical approaches to teaching and curriculum design are also explored so that anyone working with children can encourage the development of literacy skills through the Fine Arts.

INTRODUCTION

When I was a child my two passions were creating visual art and reading. I remember before I could read the words in books I would sit and examine the pictures. Page after page, I vividly remember examining the various hues used in each book, the different styles of illustration, and the various characters that were used. Although I didn't know the proper terminology or understand how to create the images or the media involved, I noticed that some books had ink drawings in them, while others had oil paintings or watercolor paintings, and others even had some pictures created out of plastercine. Some had lots of detail while others were vague and general. I wondered why some were in color and others were black and white. Later on, learning how to read the text came easily to me but even then I continued to savor the pictures in addition to the text.

While half of my spare time was devoted to reading, the other half was devoted to creating visual art. Drawing, painting, photography, ceramics, papier-mâché – it didn't matter what medium it was; as long as I was creating something I was happy. I remember creating picture with stones and twigs along the reservoir near my house, painting with food coloring on cakes and cookies, and creating papier-mâché sculptures to sell to my family. I often combined my two favorite hobbies and wrote and illustrated my own books. Although they were often simple books with similar plot lines of traveling through tunnels and discovering ghosts and skeletons, they were always fun to create and they amused my family.

Now as I watch my own three-year-old read books and create her own artwork I wonder if and how these two activities are linked. We began our adventure into reading with simple picture books, ones with lots of pictures and few words. I didn't choose these books because of their lack of written text, but because they had rich illustrations. Books we could look at, talk about, and discover new things each time we picked them up. They

ISBN 978-0-9566139-1-2

were books with bright illustrations and lots of detail, ones that we would both find appealing, such as those written by Eric Carle, Maurice Sendak and Arleen Alda. Sitting in our comfy chair, looking at beautiful artwork in books, was fun and at the time I hoped it was somehow educational. Now that I look back and reflect on the process, I know that I was actually developing her visual literacy skills. I began to wonder if this process helped both of us learn how to read. I agree with Ann Alejandro when she states that '*immersion in art can parallel and enhance immersion in text: when we read and write we use the same critical thinking and decision making skills that we use when we paint or respond to paintings*'.[1] This paper takes this idea further and explores how developing one's visual literacy may help in creating a foundation for reading development.

What is Visual Literacy?

The term 'visual literacy' has been in use since the 1960s, and yet continues to have many different definitions that are unclear and nonfunctional. According to Sinatra, the term has often been a catch-all term to mean anything that is delivered through visual sensory input.[2] It is a term that is used often in media studies to refer to the ability to understand images, or 'read between the lines', in commercials and advertisements that people are constantly being bombarded with. It can also refer to being able to understand, or interpret, visual images including ones found in books, magazines or art galleries. Visually literacy is often associated with 'visual thinking', a broader term that encompasses both understanding images as well as creating images for others to understand. In this paper I have chosen to limit the scope of visual literacy to the visual arts, particularly visual artwork found in picture books because they are easily accessible to all students. Therefore, for the purposes of this paper, I shall adopt the definition of visual literacy as 'the learned ability to interpret visual artwork and to create such artwork in order to communicate ideas and concepts'.

It is important to note that visual literacy is culturally specific. As Stokes explains, visual literacy, like language literacy, '*is culturally specific although there are universal symbols or visual images that are globally understood*'.[3] Different cultures have differing views, values and uses of the visual arts. Obviously, this creates different relationships among the arts, communication, language and literacy. For example, nations from the Pacific Northwest Coast often use the arts as a direct means of communication. They use elaborate woodcarving techniques in creating totem poles that represent Native family kinships and stories. In the past this group of people had no written language of their own and the poles and meaning of the symbols were shared orally with close friends and relatives, thus creating an important relationship between oral language and visual art. In this way, First Nations people were and continue to be visually literate directly through their artwork. However, these symbols are culturally specific as not everyone can look at a totem pole and understand the family story behind it. Specific visual literacy skills need to be learned and applied properly in order to fully understand the totem pole. Although examining different cultures and the varying interrelationships that lie between visual literacy, communication, language and literacy would be very interesting, this paper will only focus on the dominant culture in North America in order to limit the scope.

What constitutes a foundation for reading?

Learning how to read is a very complex process. Fortunately for most adults reading is an unconscious activity that has developed over time to become a seamless skill. However, for young children, learning how to read can be a daunting task that is anything but seamless. There are many factors that influence when and how one learns to read. But what kind of base does one need before he or she can understand the symbolic nature of reading? How does visual literacy contribute, if at all, to developing a foundation that serves as a springboard for reading? Interacting with visual arts, such as the art in picture books, may be a way to help young children develop a foundation upon which they can develop reading skills.

Before learning how to read, children need to acquire a foundation upon which they can build their skills. This is vital, as reading requires many various skills such as knowing how to hold a book, to paying attention to details, to recognizing, interpreting and understanding symbols. Learning how to read depends on developed language structures.[4] Children need to understand language and how it functions before they can grasp the idea of reading. They must understand the idea that oral language skills, including speaking and listening, are used to convey thoughts, feelings and ideas. Then they can begin to understand that written text is really language written down. Interestingly, children do not need to excel in their written and spoken language ability in order to be able to read; only average language skills are needed in order to develop the ability to read.[5] Researchers Gentile and Hoot also emphasize the important of language. They state that '*many children who have difficulty learning to read have*

ISBN 978-0-9566139-1-2

never grasped the link between spoken and written speech' and argue that this is why printed messages are meaningless to these student. Furthermore, their research has shown that developing visual literacy through the interactions with paintings helps to bridge the gap.[6]

Students also need to recognize the alphabet as various letters. The ability to recognize the alphabet, discerning one specific shape from one letter of the alphabet, is another factor. For example, a person who is not familiar with Chinese most likely will not be able to discern which images are letters and which ones are decorative shapes. With exposure to the images, over time, we can discern which ones are letters and which ones are shapes. Developing perceptual as well as problem-solving skills is an important step in the reading process. Students must discern, decode and then comprehend symbols when reading.

In addition to language skills, children need to become familiar with literacy in use. This includes watching others read text including books, magazines, signs and symbols. This helps children learn how printed text works, which includes understanding that print contains a message, the differences between upper and lower case letters, left to right print orientation, top-to-bottom directionality, the usage of punctuation marks, and even how to hold a book.[7]

Understanding the sequential format of stories and discussing them also helps in literacy development.[8] Not only does this help children understand how print works and develop language skills, it also aids in increasing children's motivation to become readers. Talking about stories and predicting what happens next helps to increase excitement and can motivate students to read. Motivation is an important factor that is often overlooked. If children have no desire to learn to read, forcing them will only result in frustration.

How does visual literacy help develop a foundation for reading?

1. Language and Communication

Both the visual arts and written text are a means of communication. People have been using the arts as a means of communication for thousands of years. The arts were the first and the primary form of human communication.[9] For centuries, illustrations were the primary means used to convey religious and political information to illiterate people. Today the visual arts can serve many purposes, including social and personal communication. The visual arts, whether they are direct or indirect images, ideas, or feelings, can make the ineffable communicated among people. Perhaps the ability to communicate is one of the most important and everlasting abilities that is fostered in interactions with visual images.

Practicing visual literacy can aid in developing language abilities in various ways. Children can create visual images to explore language, to understand language, and to communicate with drawings as a language. Visual images can be used to direct language use, help children practice verbal language, and to communicate in a more full, rounded manner. Bob Steele, the founder of The Drawing Network in British Columbia, has shown how children use visual arts as a language medium in his extensive research projects. He has found that from the age of about two years, children draw as though using language. Children's artwork often shows events organized in sequence (story telling); poetic and metaphoric imagery; diagrams that analyze and explain and graphic images that compare, categorize, evaluate, express feelings and emotions, describe from observation and memory, and demonstrate fantastic invention. According to Steele, many pictures drawn by children show the hallmarks of language, as children can articulate, express and communicate their thoughts, perceptions, feelings and creative imaginations through this medium.[10] Researchers Neu and Stewig have also shown that many children draw for emotional enjoyment as well as to explore expression and language.[11] These researchers also state that children use drawing as a rehearsal and scaffolding technique while learning to write, read, and comprehend language, to decode words and understand story-language patterns, and to generate and organize ideas for reading and writing.[12] So, if verbal language is a necessary precursor to literacy (as discussed previously), and practicing visual literacy through drawing can help children explore and develop their language skills, then it is possible to conclude that practicing visual literacy in such a way can help to develop a foundation for learning how to read and write.

ISBN 978-0-9566139-1-2

2. Decoding abilities (reading) / encoding abilities (writing)

Visual literacy and print literacy are two different communicative systems that are similar in the way they are processed for understanding. Both words and pictures are processed in sequential stages. According to Winn the first stages in interpreting both text and pictures are '*automatic and are beyond the willful control of the viewer*'.[13] This means that people first see lines, textures and values, and automatically separate objects from one another and their backgrounds. Once we consciously attend to these elements we analyze them in greater detail as images or letters forming words. In order to be able to read one must be able to look at print, recognize the letters and then words. Symbols are crucial in reading and interpreting text, a factor that is often overlooked when examining reading.

Instead of direct links between the words, images and sounds, in the English language words and visuals serve as substitutes for something else, otherwise known as a referent. Obviously, the degree of similarity of words and pictures to their referents differ. In other words, in some cases, pictures can more easily resemble their referents and thus they can both present similar visual stimuli, whereas words do not. However, in the case of abstract art, indirect images are often full of symbols that enable the artist to share their thoughts, feelings and ideas with others. In such a case, the image may be a combination of referents that evoke various feelings and ideas. Although both images and text serve as substitutes for their referents, extracting their meaning depends on a repertoire of learned strategies. This is true even for pictures that fall under the realistic/representational side of the continuum as the resemblance between the image and their referent may not always be apparent to all viewers.

3. Comprehension

Developing visual literacy relies on examining, interpreting and discussing various visual images. It is possible for these visual images to come from many different sources, including magazines, books, television, or advertising signs. However, for my purposes, I am focusing on the visual arts, and more specifically, visual arts found in picture books as the source of visual images. According to Barkan, artists use many forms of symbolism that encompass his or her experiences, and these symbols are then shared with viewers.[14] In order to understand the ideas and experiences that the artist wants to convey, viewers must develop and use visual literacy skills.

The first step is developing perceptual skills by interacting with the visual arts on a basic level, such as identifying a figure from its background. This is important for being able to identify written symbols on a page. Simply creating and discussing paintings can help children identify and create patterns and symbols, and become aware that images on paper are meaningful and say something.[15] Interacting with visual art on a deeper level by developing visual literacy can aid in critical-thinking skills, including skills for describing, analyzing, interpreting, and making judgments and thoughtful responses to creative works.[16] As we use our visual literacy skills and interpret what the artist has created and the meaning behind the work by examining the artwork critically, we create new understandings and develop comprehension abilities. Being able to understand symbols and comprehend them helps to develop a solid foundation for literacy.

Other factors that contribute to comprehension abilities are creativity and problem solving skills, being able to integrate concepts, and engagement with the task at hand. Students need to be able to figure out what new words mean by thinking creatively and using and applying different strategies. Participating in the visual arts and building visual literacy skills can aid to strengthen creativity, imagination, and problem-solving skills. Combining visual literacy and reading, such as examining picture books, can lead to stronger engagement and higher motivation to read the text.

4. Motivation

Many experts agree that motivation is a contributing factor in learning how to read.[17] There are many attitudinal and behavioral benefits that occur indirectly by participating in the visual arts. They can help to improve life skills such as working in groups, critical thinking skills, and self-criticism, and develop pro-social behaviors and attitudes among 'at risk' students such as by building social bonds, working with mentors, and improving self-images.[18] It is reasonable to assume that an improved overall self-concept, self-esteem and morale can lead to an overall improvement of self-expectations, goals and positive behavior, encouraging and motivating students to improve their literacy skills.

ISBN 978-0-9566139-1-2

Artists and viewers are drawn to the arts because of the expectation that the arts experience will be rewarding, pleasurable and meaningful. Art experiences for both the artist and viewer can result in a sense of satisfaction and pleasure, and add value to people's lives. Recurrent aesthetic experiences can also help develop empathy by aiding in one's capacity to participate imaginatively in the lives of others.[19] Practicing visual literacy skills by examining visual arts in picture books can be very rewarding and pleasurable. Because books are used to encourage visual literacy, students may transfer the feeling to reading in general, and thus be motivated to read even more.

5. Exposure to literacy in use

Having experience with literacy in use, or in other words, seeing other people reading and writing, is an important factor prior to learning how to read and write for oneself. This is important because it can help children understand that literacy is about deriving meaning from and communicating with others using symbols. It can stimulate interest in literacy, help children recognize the alphabet, and shows children basic skills such as how to hold a book, reading from left to right, and the various types and sources of print that are available.

Students may experience literacy in use indirectly or directly while developing visual literacy. An example of indirect experience would be a field trip to the museum with the intentions of examining the paintings. This could involve the teacher reading maps on how to get to the museum, reading the itinerary at the museum, writing notes about the children and the paintings, and reading information posted about each artist. In other cases, experience with literacy in use may be directly related to the visual literacy learning. My preferred method of teaching visual literacy is by using the artwork in picture book. Children can see and hold the book itself, examine the images from left to right, and in some books, examine at the accompanying text.

6. Transfer

While all of these skills are important to a literacy foundation, the reader might be wondering whether or not these skills can and will transfer from an arts based domain to a literacy domain. This type of transfer, also referred to as out-of-domain transfer or context transfer, refers to a skill learned in one domain and transferred to a new domain. According to Eisner, for true transfer to occur, the student must learn a skill in one area and apply it to another area on his or her own.[20] This means that the student must not be specifically taught how to accomplish the new task successfully. This true transfer of skills is very difficult to demonstrate in any study, and is what makes the concepts of transfer difficult to study and establish.

According to Philip Yenawine, it is likely that developing visual literacy can enhance the development of other meaning-making systems.[21] Visual art is a language in itself and still requires basic thought processes that are inherent in other forms of literacy. According to Fransecky and Debes, '*both visual and verbal languages involve thought processes which precede speech and writing*'.[22] They explain that language has what they call deep and surface structures. Deep structures are a result of growth, whereas surface structures are the sounds and visual symbols that communicate. Fransecky and Debes continue to explain that '*a good visual statement – a picture, painting, or film – begins with an underlying idea – a kind of deep structure – from which the communicator develops a surface structure visual presentation*'.[23] As students practice visual literacy skills, they make links between visual and verbal languages. Specifically, students will

> *begin to make linkages between verbal language composed of predicate and noun elements arranged purposely to communicate, and visual language elements (action and object elements) which are also arranged for intentional communication.*[24]

Teaching Approaches

Although being able interpret images may come easily to some, becoming truly visually literate is a learned skill. Some students may not pay attention to important details and miss subtle aspects of illustrations that are important in understanding and appreciating the story while students may focus too much on the small details and become preoccupied with aspects of illustrations at the expense of the whole.[25] Students need to learn how to examine and interpret the images in meaningful ways in order to become visually literate in a beneficial way. It is not something that develops naturally over time.

ISBN 978-0-9566139-1-2

Picture Books: The Ideal Medium

Combining written text and visual arts combines two important forms of messaging that allows for the formation of a full and well-rounded form of expression. It encourages authentic communication that allows one to communicate more than what one is capable of in words alone.

Some people argue that picture books do not promote literacy because some of them contain very few words. However, books with minimal text can still communicate vivid stories, ideas and even traditions to their readers. In addition, picture books help children learn how books work and are a fun way to introduce books to children. They are often inviting and easily capture children's attention. They teach children important pre-literacy skills, such as how to hold books, how to move from left to right, page turning skills, and even 'reading' the title, author, and page numbers. These are all important literacy skills that are often forgotten.

Picture books also tend to be better embraced by the public than visual art. Picture books are a very accessible means of artwork as they are readily available at schools and libraries, and everyone can look at them and can understand them to some degree. Picture books are, in fact, books and people can relate them easily to reading and writing which are highly valued skills. Picture books are a form of art in themselves, and they often provide inspiring works of visual art for everyone to enjoy and appreciate.

Picture books are the ideal medium for embracing visual literacy for the purposes of developing a literacy foundation. Although some believe that picture books are of no value and are simply a collection of pictures, well-crafted picture books are much more. According to Bader, a picture book is '*text, illustrations, total design; an item of manufacture and a commercial product; a social, cultural, historical document; and foremost an experience for a child. As an art form it hinges on the interdependence of pictures and words, on the simultaneous display of two facing pages, and on the drama of the turning page*'.[26] In addition, picture books tend to provide valuable experiences to readers. They hold readers' attention through powerful, vivid illustrations, they accommodate developmental differences of individual children, they provide pleasure with intellectually stimulating context, they challenge the brain by seeking patterns out of the complexity of stimulation from text and illustrations at the same time, they provoke conversation by aiding to increase children's vocabulary, and they connect experiences from home and family to stories.[27]

When examined closely, one will find that picture books are a superb method of storytelling that rely on two levels of communication. Unlike emergent reading books that follow careful design controls, picture books do not follow such controls and are not designed to solely support text comprehension. A careful relationship emerges between the text and illustrations. According to Norton (1994), there is a balance between the illustrations and the text and neither is complete without the other. In fact, four different forms of information provide different depths of information to the reader. This involves what is told to the reader through the words, what is conveyed through images, what is conveyed through a combination of the two forms, as well as the reader's personal association with the book.

An excellent example can be seen with the book *Where The Wild Things Are*, by Maurice Sendak. We can see information provided in the illustrations that is not in the text, and information in the text that is not in the illustrations. The text and the illustrations both complete each other to create a full experience. For example, the illustrations show just what kind of mischief Max is perpetrating and exactly what his wolf suit looks like. The words tell the reader that all of this happened at night and that the human figure is a boy named Max (the illustration is ambiguous if it is a boy or girl). The reader creates a personal association with the book as he or she fills in the gaps that are deliberately left for the reader. For example, Max says 'I'll eat you up!' and then he was sent to bed without supper. The text and the illustrations do not tell the reader exactly what his mom said to him. It is up to the reader to imagine the words his mom said to him. This helps the reader create personal associations with the book. These different levels of information help to create a rich and unforgettable picture book experience.

Overall the sparks that are created between the text, illustrations and the reader create a unique reading encounter that cannot be experienced in other types of books. It is what makes picture books unique and is precisely why picture books are an ideal avenue to use to develop visual literacy among children.

ISBN 978-0-9566139-1-2

Visual Thinking Strategies

While Abigail Housen doesn't use the term 'visual literacy', her questioning techniques are actually a method that directly develops students' visual literacy abilities. She uses the term 'visual thinking strategies', or VTS.[28] Housen's strategies are questioning techniques designed to observe, evaluate, synthesize, justify and speculate in combination with visual art. This is one of the most involved and thought-out studies that shows how visual literacy can directly affect critical thinking skills and reading levels. Housen also shows transfer between the fields of visual literacy and reading, an important factor that is difficult to show.

Visual Thinking Strategy (VTS) utilizes specific open-ended questioning techniques that encourage all students to participate in discussions and use a combination of visual skills and their own prior knowledge to come to their own conclusions about a piece of artwork. The three founding questions are: What's going on in this picture? What do you see that makes you say that? And, what else can you find? Teachers are encouraged to listen carefully to student responses, physically point to what the student mentions in the image, accept all student comments, paraphrase comments to demonstrate proper sentence construction and vocabulary, and finally, link answers together to show how ideas are linked and build upon each other. The goal is to actively engage the viewer, developing connections to what he or she sees, and come up with critical, thoughtful responses. This process takes time and teachers must be able to be patient with students and provide them with the opportunities to examine the artwork, think about it, and fully participate in discussions.

While participating in VTS, students willingly practice their verbal skills, increase their vocabulary base, develop supported interpretations of the artwork, develop critical thinking skills, and build confidence in themselves. All of these are important in developing and improving literacy because while reading, students need to be able to draw conclusions and make inferences supported from the text, use critical thinking skills to think about what they are reading and apply it to different situations, and be able to understand and approach new vocabulary confidently. While writing, students need to support their arguments with examples and evidence, use and take risks with new vocabulary, critically think about what they have written, and be confident in their work. In addition, a good oral language base that is practiced in VTS can only help students with both reading and writing.

The Whole Book Approach

The Eric Carle Museum in Massachusetts is a unique institution as it fully embraces teaching visual literacy. Obviously it is dedicated to the work of Eric Carle and thus focuses on both his visual art and his picture books. When examining visual artwork with children the museum uses VTS as described above. However, when examining picture books the museum uses the Whole Book Approach.

The Whole Book Approach (WBA), based on a dialogic reading strategy, focuses on creating full reading experiences.[29] This means that story time facilitators include discussions about how the pictures, book design, and production elements of the picture book work together to create an artistic whole. Students are encouraged to ask questions, share reactions, and participate in conversations while the story is being read, therefore creating an experience of reading with children rather than reading to children. Pausing and conversing with listeners allows them time for reflection, clarification, and expansion of ideas. Children become engaged in careful observation, listening, and critical and creative thinking.

See, Think, Wonder

Richards and Anderson developed another approach that was designed to help develop visual literacy skills. The See, Think, Wonder (STW) model refers to the questions 'What do I See? What do I Think? What do I Wonder?' and was designed to stimulate emergent reader's imaginations, interest, and curiosity through interacting with others.[30] It encourages students to examine nuances of illustrations that supply pivotal information about stories and appreciate illustrations as unified compositions that convey important concepts, such as reality or fantasy, sadness or joy, and the main idea of stories.

It should be noted that although the STW model is basic, it can be extended by encouraging students to engage in multiple literacy experiences such as drawing and painting illustrations they think should have been included in a storybook. The ways different illustrators use the elements and principles of art can also be compared and contrasted, such as the black and white values used in Selznick's The *Invention of Hugo Cabret* to the primary colors in Dr. Seuss' *Cat in the Hat*. Multicultural picture books can be examined to determine if the illustrations

ISBN 978-0-9566139-1-2

depict accurate representations of people and customs. Finally, various versions of the same picture book by different authors and illustrators can be compared and contrasted.

A Suggested Approach for Educators

Although there are many forms of visual art available that may be used for the purposes of expanding visual literacy, I suggest that educators use picture books as the primary medium. Picture books contain many types of quality visual images created with a variety of media and are easily affordable and accessible. As previously mentioned, picture books fully support the development of a foundation for reading as they provide opportunities to expand visual literacy and opportunities for students to experience literacy in use concurrently.

However, the reader may wonder what strategy to use. VTS is a strategy designed for use with visual artwork such as paintings and drawings. VTS questioning techniques do fully engage the viewer and focuses on visual literacy. However, since it is a technique that focuses on visual arts in gallery and museum settings, it does not take into consideration picture books as a whole and their unique importance and contribution to visual literacy. Unlike VTS, WBA does focus on picture books as a whole and the importance of the text and images coming together to form a unified whole. Examining the book as a whole does engage readers and helps them think critically about the book as a whole. However, it does not directly focus on expanding visual literacy. Finally, the STW approach does encourage students to stop, look and think about the illustrations in a book, it does not encourage students to interpret the illustrations and examine the nuances and their importance to the story and text enough. Therefore, in order to fully expand students' visual literacy experiences I suggest carefully combining the three approaches previously discussed.

Combining the questioning techniques will engage all viewers. Asking all of the questions, namely, what do I see/what's going on in this picture, what do I think/what do you see that makes you say that, and what do I wonder/what else can you find, will help ensure that all students understand the questions, think about them in multiple ways, and examine the images from various viewpoints. Educators who listen, paraphrase, point out details in the picture books, and link ideas to student experiences will help to create interest and help students understand the visual symbols they are decoding. Learning will become authentic and relevant for the students. Combining this with the WBA will help students understand the importance of books and become critically engaged with the book as a whole, resulting in a complete visual literacy experience designed to strengthen the foundation needed for literacy.

CONCLUSION

Connecting ideas from the opposite fields of Fine Arts and Language Arts can help students build a literacy foundation. Picture books are an ideal avenue for developing visual literacy to prepare students for literacy learning. Visual literacy can help develop language and communication skills, the ability to encode and decode symbols, aid in comprehension, and motivate students. Moreover, focusing on visual literacy using picture books provides students with direct experience with literacy. This creates a foundation consisting of a basic understanding of language, an ability to decode and encode symbols, strong perceptual and comprehension skills, and motivation to learn. It also provides students with a familiarity with literacy in use and general knowledge about the importance of literacy. All of these factors play an important role in creating a strong foundation that serves as a springboard for reading and writing success.

REFERENCES

[1] Alejandro, A. Handbook of Research on Teaching Through the Communicative and Visual Arts. (1997) New Jersey, U.S.A.: Simon and Schuster, p. 795.

[2] Sinatra, R. *Visual literacy connections to thinking, reading, and writing.* (1986) Springfield, Ill. U.S.A.: C.C. Thomas, p. 46.

ISBN 978-0-9566139-1-2

[3] Stokes, Suzanne. "*Visual Literacy in Teaching and Learning: A Literature Perspective*", Electronic Journal for the Integration of Technology in Education, (2001), College of Education, Idaho State University.
[4] Nevills, Pamela and Wolfe, Patricia A. *Building the Reading Brain PreK – 3*. (2009) Thousand Oaks, California: Corwin Press.
[5] Nevills, Pamela and Wolfe, Patricia A. *Building the Reading Brain PreK – 3*. (2009) Thousand Oaks, California: Corwin Press.
[6] Gentile and Hoot, *The Reading Teacher*, Vol 36, No. 4.. (1983), p. 436-439.
[7] McDonald, N., and Fisher, D. *Teaching literacy through the arts.* (2006) New York : Guilford Press, p 39.
[8] McDonald, N., and Fisher, D. *Teaching literacy through the arts.* (2006) New York : Guilford Press, p 39.
[9] Cornett, Claudia E. *Creating Meaning Through Literature and the Arts, Second Edition*. (2003) Upper Saddle River, New Jersey: Pearson Education Inc.
[10] Steele, B. *How Daily Drawing Helps Children Achieve Literacy*. (2004) The Drawing Network Handout.
[11] Bates, Catherine. *The Resource-Teacher Project Background*, (2006) http ://www.nald.ca/canorg/ cclow/doc/Drawing/4.htm
[12] Bates, Catherine. *The Resource-Teacher Project Background*, (2006) http ://www.nald.ca/canorg/ cclow/doc/Drawing/4.htm
[13] Winn, W.D. 1990 "A Theoretical Framework for Research on Learning from Graphics." *International Journal of Educational Psychology* 73, (1990) p 57-64.
[14] Barkan, Manuel. *A Foundation for Art Education*. (1955) New York, NY: The Ronald Press Company.
[15] Gentile and Hoot, *The Reading Teacher*, Vol 36, No. 4.. (1983), p. 436-439.
[16] The British Columbia Ministry of Education, *The Primary Program: A Framework for Teaching*. (2010) http://www.bced.gov.bc.ca/primary_program/primary_prog.pdf
[17] Nevills, Pamela and Wolfe, Patricia A. *Building the Reading Brain PreK – 3*. (2009) Thousand Oaks, California: Corwin Press.
[18] McCarthy, K., Elizabeth O., Laura Zakaras and Arthur Brooks. *Gifts of the Muse: Reframing the Debate About the Benefits of the Arts*. (2005) Santa Monica, CA: Rand Corporation.
[19] McCarthy, K., Elizabeth O., Laura Zakaras and Arthur Brooks. *Gifts of the Muse: Reframing the Debate About the Benefits of the Arts*. (2005) Santa Monica, CA: Rand Corporation.
[20] Eisner, E. W. *The Arts and the Creation of Mind*. (2002) Harrisonberg, Virginia: Yale University Press/New Haven & London. p. 220.
[21] Yenawine, Philip, James Flood, Shirley Brice Heath, and Diane Lapp. *Handbook of research on teaching literacy through the communicative and visual arts: Thoughts on Visual Literacy*. (1997) New York: Macmillan Library Reference USA; London: Prentice Hall International, p. 846.
[22] Fransecky, Roger B. and John L. Debes. *Visual Literacy – A Way To Learn, A Way To Teach*. (1972) Washington, D.C.: Association for Educational Communications and Technology, p. 7.
[23] Fransecky, Roger B. and John L. Debes. *Visual Literacy – A Way To Learn, A Way To Teach*. (1972) Washington, D.C.: Association for Educational Communications and Technology, p. 8.
[24] Fransecky, Roger B. and John L. Debes. *Visual Literacy – A Way To Learn, A Way To Teach*. (1972) Washington, D.C.: Association for Educational Communications and Technology, p. 7.
[25] Richards, Janet C. and Anderson Nancy A. "*What Do I See? What Do I Think? What Do I Wonder*", The Reading Teacher, Vol. 56, 2003.
[26] Perkins, Arizpe, and Styles. *Intelligent Eye: Learning to Think by Looking at Art.* Children Reading Pictures. (2003) New York, NY: RoutledgeFalmer, p. 21.
[27] Jalongo, M. R. Young children and picture books (2004), Washington, DC: National Association for the Education of Young Children.
[28] Housen, A. *What is VTS?* (2010) http://www.vtshome.org/pages/what-is-vts
[29] The Eric Carle Museum, "*The Whole Book Approach", (2010)* http://www.carlemuseum.org/downloads/Whole%20Book%20Approach%20Final.pdf
[30] Richards, Janet C. and Anderson Nancy A. "*What Do I See? What Do I Think? What Do I Wonder*", The Reading Teacher, Vol. 56, 2003.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

INGREDIENTS OF IDENTITY: FOOD, COMMUNITY AND INTERDISCIPLINARY RESEARCH

May ROSENTHAL SLOAN

PhD Student, University of Glasgow, United Kingdom
mayrosenthal@yahoo.co.uk

Keywords: Food, Identity, Ethnicity, Community, History, Geography, Anthropology

ABSTRACT

Margins can be spaces of creative overflow, where ideas formed within mainstream research methods break into new territory. This type of space is reflected in society, where those considered as existing in the margins of society are often people whose lives and work are particularly worth understanding.

The idea of academic and societal margins appeals enormously and has helped me to define the research that I am undertaking. This paper will serve as an introduction to this work, which looks at the relationship between food culture and identity building, in a set of five streets in Dalston (London) and Flatbush (Brooklyn, NYC). The two locales, chosen precisely for their often marginal status, are traditionally poor areas, made up of various immigrant communities, who live alongside working and middle-class natives of the two cities. Both have (among a large range of other peoples) notable Caribbean populations, strong legacies of Jewish settlement, and more recently a build-up of artists, attracted by distinctive cultures and relatively low rents. My research is concerned with what can be learned about how such different groups relate to one another, by looking at attitudes toward food.

Methodologically, I feel that I exist in various academic margins. My work sits perched at the edge of Social and Cultural History, comfortably dipping its toes into Anthropological and Geographical pools of research. Oral History is a primary methodology, and the aim to share the control and ownership of the work with those whose identities I seek to analyse, places it in a fairly unique space between academy and community. I hope to reflect such positive marginal status in the presentation of my work, utilising language both academically rigorous and popularly accessible, and I feel strongly that such work can and should aim to be valued on these different levels.

INTRODUCTION

One of the only things shared the world over, often with great cultural significance as well as fundamental physical need, is the consumption of food. It is my belief, and one which has recently enjoyed (and continues to enjoy) growth in the academic world, that an analysis of food habits and attitudes can illustrate a great deal about an individual, a family, or a community. I also believe that looking at food is a valuable way of trying to understand cities, particularly those made up of diverse ethnic groups, whose culinary histories may differ, but whose contemporary space is side by side with one another. My work, based in History, but with distinct influences from schools of Geography and Anthropology, seeks to explore the relationship between food cultures and ethnic identity in two very different, but both extremely diverse, urban environments, one each side of the Atlantic. Exploring individual histories of those who buy, sell and eat food in these environments, and using the physical present as a series of historical sources, I hope to turn a unique analysis of the past, into a better understanding of how communities experience today and tomorrow. The following serves to introduce the research which I am carrying out for my PhD, which is in its early stages.

ISBN 978-0-9566139-1-2

METHODOLOGY

Methodological choices are particularly important when dealing with issues of community and with the lives of real and living people. An approach which combines historical, geographical and anthropological methods will be utilised to achieve the best results.

History (oral and otherwise)

A firm base of oral history seems to me, the most appropriate way of exploring the questions I will be asking. I am extremely keen that the work produced should be genuinely reflective of the communities it seeks to understand. As with all historical research, this is a project shaped by interpretation, and I feel that on a very basic level, the use of oral history reminds the historian, and the reader, that such research is never without a level of partiality, and research outcomes are necessarily informed by the process through which they come to be, both by the researcher, but in the case of oral history, also through the participants.

While recruiting interviewees, it will be my aim to talk to a good representative group from both Dalston and Flatbush. However this is a wide and thus challenging combination of communities, within which it is likely that I will have varied levels of ease in terms of recruitment. It is important that if some are more willing than others to be interviewed, this is reflected in the work, for people's reactions to me as an interviewer are indicative of their relationship with their environments, and as such are of significance to the research itself. Applying oral history theory, as well as wider historical research, I believe that the process of recruitment can and should be one which sheds light on and benefits the wider research of which it is a part.

When interviewing participants, I will take a life history approach avoiding specific questioning. In doing this, I hope that the control of the interview will to some extent be shared between myself and the interviewee, allowing them to establish a narrative that is most representative of their lives. I have the luxury in terms of my subject matter, of being able to take such an approach as it is very rare for anybody discussing their own history to leave out their experiences of eating. Its universal nature is such that it will add interest should any participant fail to mention food naturally. Thus I will be able to allow the interviews to flow with minimal guidance, and will hopefully establish the level and nature of importance that food has in the lives of my interviewees. Full transcriptions will be made of each interview, allowing me to analyse with the fullest possible sensitivity the information imparted, as well as giving interviewees something tangible to take away from the process.

Space, Place and Anthropological Influence

Given the desire to discuss distinctive ethnic mixes and, rather than looking at specific groups, discuss the relationships between different peoples, it has become increasingly important to limit my research in other ways. The concept of experiential space between ethnic groups in urban settings has seemed to encourage an organic growth of the spatial in my research, and led me to impose strict limits upon my areas of selection based on space. I believe this is both a sensible practical step, but also poses interesting questions of its own which have led me to look at my work in new ways. The selection process of five streets in London and New York, upon which to base my research, has led me to compare the physical spaces in which food is bought, sold, and eaten, and question what effect local geography might have upon these processes. Mapping the landscape of commercial food in the two areas, both past and present, will allow exploration both of how space is used, and of what culinary changes have taken place. Throughout this, the trans-Atlantic element will come into focus, each city encouraging a stronger analysis of the other. Reducing the spatial focus, I am interested in how small, individual spaces are utilised by different groups, in different areas and what can be learned about communities and their history by looking carefully at buildings, market stalls, and the objects; and of course, the food, housed within them. Sensitive and prolonged observation of the two physical areas in question, and the habits of those within them, will I believe add a great deal to the historical research undertaken. The analysis of material gathered through the combination of methods discussed will allow for the best possible understanding of the locales I have chosen to talk about, and the identities of those within them.

ISBN 978-0-9566139-1-2

DISCUSSION

Food, in the academic sense, appears to be striding confidently out of the margins, having built in production and following over the last forty years. Writing in 2001 in the preface to her book *Hungering for America, Italian, Irish and Jewish Foodways in the Age of Migration*, Hasia Diner describes the nonplussed responses of other academics to her work upon food: '*Some just think it a strange subject for an historian. Treaties and wars, strikes and elections, rallies for equal rights, protests against discrimination and exploitation, these are the stuff of history. Food seems ephemeral, the subject for cookbooks*'.[1] Ten years later, it feels on both sides of the Atlantic as though food as an academic subject has stepped out of the margins, and is enjoying the start of its life as an accepted mainstream area of study. Diner has a clear approach defined along fairly traditional lines of historical scholarship and this works well for her. Particularly interesting is her analysis of Irish experiences and the influences of colonialism, where she argues that the lack of a discernible Irish food culture in the United States was a result of reluctance to emulate the elite 'others' of Ireland, the English colonisers. However, I believe that the most dynamic research can use imaginative historical analysis such as this, as a basis for direct understanding of contemporary urban cultures. Josée Johnston has asserted that '*food by its very nature flows across disciplinary boundaries*' and that embracing this helps in a quest to build a picture of amongst other things, how food relates to ethnicity, gender and class.[2] Such a statement is, in my opinion, absolutely appropriate. Helena Tuomainen, whose research into Ghanaians in London draws on Diner's treatment of Irish colonialism, to discuss contemporary culinary experience, illuminated by a historical understanding of British colonialism in Ghana, demonstrates with particular skill and sensitivity the potential for interdisciplinary research into food and culture.[3] What I aim to do is influenced by Tuomainen's approach, as well as by historians of American food culture such as Diner and Marcie Ferris-Cohen, whose use of oral histories and recipes within academic text is interesting.[4] However, what all of these scholars do is to highlight the interesting dynamics surrounding minority groups in specific place-defined cultures and in each case, there is a sense that the effects of these minority groups upon those cultures would merit further investigation. Stated differently, the effect of the marginal upon the mainstream deserves more attention in terms of culinary culture. Donna Gabaccia and Panikos Panayi, writing respectively about North American and British food culture redress this somewhat, both avoiding the specific ethnic focus, although both of their works differ from what I aim to do in that (although based on city case studies in Gabaccia's case) they deal with food culture nationally.[5] In my own research, rather than looking an ethnic group, I wish to set my limits geographically and use spatial boundaries. In doing this, I hope to develop an understanding of different groups, whose common ground is the physical space in which they sell or buy food, and how these groups do or do not relate to one another culinarily. I believe that in such settings, food will serve as an interesting and appropriate social lens.

Observations of what is available for purchase in the areas in question, and in what form, encourage an insight into the areas themselves, and the communities which they house. To observe what foods can be obtained (in varied forms) from raw produce, through prepared snack foods and takeaways to restaurant meals, is a process in the analysis of the area in which research is being carried out. Why, for instance might the food of one ethnic group be available in takeaway form, to be eaten on the street, while that of another is more usually sold to be eaten sitting down at a table? Such questions allow analytical excursions into different cultures of eating which in turn allow for the everyday relations between groups, past and present, to be better understood. In terms of comparative analysis between the two locations (which will be discussed in more detail later on), the question of physical landscape is important; and I am concerned with whether the very different geographies of Dalston and Flatbush have in any way effected what is traded and eaten; or whether culinary cultures (particularly those of incoming ethnic minority groups) endure or are altered in similar ways, regardless of the physical environments in which foods can be sold. Particularly, the question of casual eating, and the implications of physical space for such practices is of interest, and I am keen to explore the effect of landscape on different street foods, takeaways and the question of where and how these items are consumed.

Looking directly at the foods which can be purchased is logically followed by the question of who is doing, and has done, the purchasing; and indeed, who they purchase from. The question of how one's cultural (and especially ethnic) identity effects and is effected by the foods one chooses to eat is absolutely central to the research which I am carrying out. I believe that an analysis of food choices, and attitudes towards food[6] is an invaluable tool in understanding ethnically complex communities. When choice is wide, choices can be loaded with meaning, and it seems likely that the level of willingness to take on culinary influence from immediate neighbours could be an indicator of relations with those neighbours. It is my belief that such dynamics can be

ISBN 978-0-9566139-1-2

better understood through parallel analyses of contemporary practice, and historical experience. Seen another way, the present is used as a historical source to explore the past. At the same time, an effort to understand the past, allows us to analyse our current situation.

Moving on from the questioning of who buys what food, is the question of who sells what food, and how this effects the interaction of different peoples. During research for my undergraduate dissertation in Ridley Road Market, one of the London locations for my doctoral work, interviewing market stall holders revealed a somewhat unexpected relationship, whereby many of the white British Londoners who had previously sold apples, potatoes and other relatively familiar fruit and vegetables, now had often thriving businesses selling yam, cassava, mangoes and other Caribbean produce to the large local West Indian community. I found also that gender seemed to be a particularly important issue in terms of whether the sellers of these items were effected in everyday eating terms by this change. While the number of participants interviewed was far too small to make any meaningful judgement, it seemed that the women who operated such stalls in the market interacted with their Caribbean customers on a level which was concerned with the preparation of food and as a result, began to incorporate Caribbean cooking practices and ingredients into their own diets and those of their families. The men I interviewed who ran similar stalls, did not appear to take on these influences in quite the same way, as they were not responsible for the preparation of food within their families, although there were examples of inter-ethnic relationships based on other things: for example, one respondent was keen to offer what he saw as an education in British selling to his new Iraqi neighbour in the market. Being given the opportunity to test such ideas with a bigger and more complex project is particularly satisfying, and the further layer of a trans-Atlantic comparison adds to this.

A research issue which provides excitement and difficulty in fairly equal measures is that of attempting to represent those people often considered to be on the margins of society. The food industries of London and New York (as with most large cities) are natural areas for those who are economically, socially and legally vulnerable. The lives of these people, and the effect they have upon the cities in question are important issues, and it seems absolutely right that those often silent should be given a voice through this kind of research. However, this will not be easy, as those in vulnerable positions are often the most difficult to gain access to. Sensitivity of approach and persistence will be called for, if I am to gain the trust of those working in these traditionally poor, ethnically varied areas. However, any difficulty or indeed failure to achieve these ends will also be valuable grounds for analysis, in that the responses of potential participants to me, as an academic, will be reflective of their perceived place in the environments at which I am looking.

In terms of output, I feel strongly that what is produced should not only be reflective of, but also accessible and relevant to, those whose experiences are recorded. It would seem inappropriate to write from a 'superior' height, in a way which excluded the very people who had made the work possible. I do not wish to write in the way that John Brewer describes as 'prospect history' and believe that there is no reason for my research to be hidden behind convoluted and overly-technical language.[7] By writing clearly and simply, and including photographs and recipes (both as historical sources, but also to maintain variety and interest throughout the writing), I believe this work should be accessible and interesting to people on a number of different levels.

CONCLUSION

In looking at the vast variety of food available in areas such as Dalston and Flatbush, at the people who buy and sell that food, and the choices that they make, I aim to come to a contemporary and historical understanding of cultural issues at play in urban environments. As previously stated, when choice is wide, choices can be loaded; and by observing and listening to the habits and attitudes of inhabitants, I hope to cast some light upon how ethnic identities are built and developed when different groups live, work and eat in the same (or at least joined) spaces. Limiting the research with geographical boundaries, exploring the physicality of areas, and combining this with strong historical research into the experiences of inhabitant individuals and groups, I hope to achieve a unique method of analysis which will allow a better understanding of how diverse urban societies come to being, and add something new to the scholarship of multi-ethnic cities.

ISBN 978-0-9566139-1-2

[1] Diner, Hasia *Hungering for America, Italian, Irish and Jewish Foodways in the Age of Migration* (2001) Cambridge: Harvard University Press, p.xv.
[2] Johnston, Josée 'Struggles for the Up and Coming: Challenges Facing New Scholars and Food Scholarship' in *Food, Culture and Society* (2008, volume 11, number 3).
[3] Tuomainen, Helena 'Ethnic Identity, (Post)Colonialism and Foodways: Ghanaians in London' in *Food, Culture and Society* (2009, volume 12, number 4).
[4] Diner, *Hungering for America*; Ferris, Marcie Cohen, *Matzoh Ball Gumbo, Culinary Tales of the Jewish South* (2005) Chapel Hill: University of North Carolina Press.
[5] Gabaccia, Donna *We Are What We Eat, Ethnic Food and the Making of Americans* (1998) Cambridge: Harvard University Press; Panayi, Panikos *Spicing up Britain, the Multicultural History of British Food* (2008) London: Reaktion.
[6] It is worth noting that these are not the same thing, and that perceived notions of what is eaten do not always correspond directly with actual consumption.
[7] Brewer, John 'Microhistory and the Histories of Everyday Life' in *Cultural and Social History* (2010, volume 7, issue 1) p.98.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

THE PALM HOUSE KEW, 1844-48, a case study on scientific experimentation and cross-disciplinary working methods in 19th Century architecture

Henrik SCHOENEFELDT

PhD Student, University of Cambridge, UK
hos21@cam.ac.uk

Keywords: Environmental Design, History of Science, Cross-disciplinary working methods

ABSTRACT

The paper discusses the significance of scientific experimentation and cross-disciplinary working methods in the design development of the Palm House at Kew, focusing on the issue of lighting as one of the main environmental design criteria. It shows how the design team, which not only comprised architects, engineers and manufacturers but also gardeners, botanists and a physicist, acting as scientific advisors and design critics in the design development process, collaborated in order to achieve the specific lighting conditions required for cultivating tropical plants in England. The project correspondence reveals that the highly transparent glazing originally projected by the engineer and architect, was rejected for transmitting excessive levels of solar radiation found to injure the foliage of plants. The experimental physicist Robert Hunt was subsequently hired to develop a special type of solar control glass, tinted to moderate the radiation. The purpose of his experiments, apart from validating that the glazing effectively reduced the transmission of solar radiation, was to ascertain that the filtered light did not produce any unwanted side effects on the natural growth of plants. The final section discusses the performance history of Hunt's lighting control, using reports and commentaries published in contemporary journals. It illustrates that Kew's authorities, starting in the late nineteenth century, were confronted with increasing difficulties with maintaining sufficient levels of light for the growth of the tropical plants, requested a scientific re-evaluation of the performance of the Palm house glazing using more advanced scientific methods and finally replaced the coloured panes with clear glass. Apart from doubts about the validity of the findings of Hunt's original experiments, the lighting was severely reduced by the discolouring of the glass under solar influence, the inception of the ambient sunlight by atmospheric pollution and deposits of black soot.

INTRODUCTION

The Palm House at the Royal Botanic Gardens at Kew, designed and constructed between 1844 and 1848, represents an early example of a cross-disciplinary and collaborative design approach, adopted in order to achieve an artificial environment for the preservation of Kew's valuable tropical plant collection, which could not be sustained in England's temperate climate. The Palm House was the outcome of a distinctively horticultural approach to building design development, which was distinct from the formal architectural doctrines of contemporary architecture. Eighteenth and nineteenth century horticultural literature illustrates that the requirements of horticulture provided the impetus for the development of new design criteria, theories and working methods, which were adopted by horticulturists to meet the specific functional, largely environmental requirements of plant cultivation.[1] The emergence of the glasshouse as a distinct building type reveals a process of innovation that was driven by a conception of architecture as primarily a design discipline concerned with the cultivation and preservation of plants.

While existing studies on the Palm House are dominated by a longstanding debate about the respective influence of the engineer and architect on its overall design,[2] the original correspondence[3] reveals that the Palm House was

ISBN 978-0-9566139-1-2

the outcome of a more extensive cross-disciplinary and collaborative design process, involving among others, physicists, botanists, practical gardeners, structural engineers, manufacturers and contractors.[4] The impetus for the adoption of these working methods was the specialist knowledge and skills that were required to resolve the largely scientific, technical and environmental questions underlying the design of the Palm House, such as the provision of ideal lighting conditions and the manufacture of ideal indoor climates, tailored to the needs of tropical plants. Science, by providing the fundamental knowledge and tools required for the formulation and application of environmental design principles and criteria, was an integral part of the ‚horticultural design' approach. The use of science, however, was not limited to building physics but also provided horticulturists with the means to acquire a more detailed understanding of the nature of heat, light, solar radiation and relative humidity and its respective influence on plant vigour.[5] At Kew, the physicist Robert Hunt conducted scientific experiments on the influence of light and solar radiation on the vigour of tropical plants and studied the optical properties of different coloured glasses. The research provided the scientific foundations for the development of a new type of glass, intended to protect light sensitive plants from exposure to excessive quantities of solar radiation.[6] The evolution of the glasshouse, both conceptually and technically, was thereby closely connected with the development of the physical and biological sciences in the nineteenth century and formed a junction between the history of architecture and the history and philosophy of science.

This paper investigates the role of scientific experimentation in the design development of the Palm House, using Hunt's work on ‚lighting,' which represented one of the chief environmental criteria underlying the design development, as an example. In addition it investigates how the actual performance of the use of tinted glass as a solar control strategy and was assessed by contemporaries and shows how, in order to adapt the building to a drastic decline in the available sunlight, the Kew authorities were forced to abandon the green glass.

'It will be so clear that you won't see that glass is there'[7]

In a letter to Alexander Milne of the Department of Woods and Forests, dated November 23rd 1845, the iron-contractor Richard Turner proposed to replace the clear glass, which had been specified for the glazing of the Palm House, with coloured glass and suggested that he procure samples with 'violet tinge' or 'azure hue' from glassworks in Antwerp and Bohemia.[8] His objective, he wrote, was to temper the intense sunlight transmitted through clear glass in order to protect humans from the discomfort of intense glare and to shield the plants from exposure to excessive sunlight, responsible for injuring tender foliage and flowers.[9] Kew's curator John Smith, who had extensive practical experience with the performance and management of the glasshouses at Kew and had acted as a consultant and critic on various aspects of the design, first brought the issue of lighting control in glasshouses to Turner's attention.[10] Smith, however, rejected the use of clear glass for being unsuitable for the glazing of the Palm House. Referring to his personal experience with the use of sheet glass in various glasshouses at Kew, he warned Turner that clear sheet glass injured light sensitive tropical plants, such as orchids and ferns, due to the high levels of sunlight that entered the building completely unchecked.[11] The extreme transparency therefore had to be retrospectively reduced by means of canvas blinds, which were used to shield the plants from the intense sunlight during parts of the summer. Arguing that the use of shades was not practical for the Palm House, he proposed to protect the plants by means of green glass. That green glass effectively protected plants, Smith reported, had been observed by gardeners in the old glasshouses at Kew, which, prior to the introduction of clear glass, were glazed with a type of crown glass which was named 'Stourbridge Green' and had a green tint.[12] Turner's proposal, however, represented a serious re-appraisal of the original design objectives. The project correspondence illustrates that an important objective underlying the design was to provide the plants with access to the highest possible levels of sunlight, in particular during the winter period when sunlight was very limited.[13] It shows that the lighting of the Palm House was subject of extensive conversations between Lindley, Hooker, Smith, Burton and Turner, which included, among others, discussion about how the lighting can be optimized through the form, orientation and construction of the building.[14] From a very early stage in the project, the project architect Decimus Burton insisted the glazed elevations should be designed specifically to optimize the admission of natural light. He rejected the inclusion of any architectural ornament, stressing that it would seriously compromise the function of the glass envelope as a light transmitter.[15] To achieve a glasshouse envelope of great transparency Turner and Burton embraced the most recent iron and glass technology.[16] The project correspondence illustrates that Turner, in collaboration with the iron founders Messrs Malins, went to great length to develop a very light wrought-iron structure to replace the considerably heavier cast-iron ribs that had been specified in the original working drawings and contract.[17] These wrought-iron ribs, due to their

ISBN 978-0-9566139-1-2

minimal cross-section, considerably reduced the portion of sunlight that was obstructed by the opaque structural members required for its construction. Before considering the use of coloured glass, Turner and Burton's objective behind the use of sheet glass as an alternative to the more commonly used crown glass, was to boost the transparency of the glasshouse envelope by exploiting the large panes that had become available in sheet glass.[18] While standard sizes of the crown glass used by Joseph Paxton in the early 1830's measured three by six inches, the panes used at Kew measured three feet by nine and a half inches.[19] To use these large panes in their full size in a curvilinear envelope the, panes were bent into a curve by the glass manufacturers to follow the curve of the iron sash bars.[20] These panes were assembled into a continuous curvilinear glass skin, by means of slender wrought-iron sash bars. Furthermore Turner thought the distinct clearness of sheet glass as desirable for horticultural purposes, saying that his aspiration was to use 'It will be so clear that you won't see that glass is there'[21] However, Smith and Turner were not the first and only people who had criticized the proposed design from the perspective of solar control. W J. Hooker, the director of Kew, had already warned in April 1844 by the botanist John Lindley, that the quantity of solar heat admitted into the building would be extremely high due to the form and orientation of the glazing.[22] Lindley was seconded by the physicist Robert Hunt, who stressed the injurious levels of solar radiation to which the plants were subjected, was not only due to the clearness of the glass but also a result of the form and orientation of glazed skin.[23] Although Hunt and Hooker did not give any detailed explanations, the project correspondence show that the curvilinear glass surfaces were originally adopted by Turner and Burton with the intention of boosting, not moderating, the levels of direct sunlight transmitted through the glazing, in particular during the winter period when sunlight was limited.[24] The reasoning underlying the adoption of the curvilinear glass envelope at Kew builds on earlier theories, formulated, among others, by the horticulturists Andrew T. Knight, George Mackenzie and John Loudon.[25] Since the proportion of the available sunlight that was transmitted through a pane of glass was dependent on the respective angle of solar incident on a pane of glass, these men argued, the adoption of a curvilinear glasshouse would ensure that the aspect of certain parts of the glazing was at most times parallel to the direction of the sunlight and thereby boosted the ratio of the available sunlight that was admitted across all seasons.[26]In addition the shallow and elongated floor plan and the stepped cross section was intended to keep plants of different sizes as close as possible to the glazed surfaces in order to ensure that the plants get their best exposure to the full influence of direct sunlight, which, as various accounts in the *Magazine of Botany* illustrate, was believed to diminish with increasing distance from the glazing due to prismatic refractions caused by the glass.[27] The correspondence reveals that the concerns raised by various critics of the design was taken very seriously by the director and the Commissioners of Woods and Forests, who subsequently commissioned a scientific inquiry in order to produce the lighting conditions that were required to sustain the vigour of the plants. Hooker, Lindley, Smith and Hunt subsequently engaged in the development of a special type of coloured glass to moderate the transmission of injurious levels of sunlight and solar heat. This inquiry, however, represents an effort to correct earlier design decisions but also highlights the contradicting aspirations underlying the urge towards "transparency" in the design of the building envelope. The following sections of this paper will illustrate that the re-appraisal of the original objectives underlying the proposed glazing system exemplifies a shift from a preoccupation with transparency towards an inquiry into ways of controlling and moderating the transmission solar light and heat. Aiming at mitigating the unwanted side effects of the unprecedented levels of transparency that had been accomplished through the most recent iron and glass technology, the design history of the Palm House exposed a conundrum of the development of highly transparent structures in horticulture.

Implications of sheet glass

Before proceeding, it is important to understand that the introduction of sheet glass had serious implications for the design development of glasshouses in general, not only for the Palm House. Extensive discussions conducted by gardeners on the letters page of the *Gardener's Chronicle* between 1847 and 1849 reveal that the scorching of plants grown under sheet glass had not only been observed in various glasshouse at Kew but was a very wide spread problem and presented the designers and operators of glasshouses with unprecedented environmental design issues. Up to this point, the experience of gardeners with the management of heat and light in glasshouses was based almost exclusively on houses glazed with crown glass, while introduction of sheet glass had been a very recent innovation in glasshouse construction. The method for manufacturing sheet glass was first introduced to Britain in the early 1830's by the Chance Brothers of Birmingham and was first introduced into glasshouse construction by Joseph Paxton in the late 1830's.[28] Commentaries in the contemporary horticultural literature reveal that the actual causes of the scorching of plants grown under sheet glass and the effect of sheet

ISBN 978-0-9566139-1-2

glass on the glasshouse environment in general, was subject to detailed field observations. According to eyewitness accounts in the *Gardener's chronicle*, the scorching of foliage under sheet glass was caused by irregularities in sheet glass, which, by acting like meniscus lenses, exposed plants to rays of concentrated sunlight.[29] When placed in the foci of these lenses, it was observed, the light left brown or white marks on the flowers and foliage of delicate plants.[30] The degree of injury to plants varied from individual spots on leaves and flowers to continuous lines on the foliage which followed closely the path of the concentrated beam.[31] The irregularities were reported to be a by-product of the flattening process used in the manufacture of sheet glass.[32] According to *Glass and Glass Manufacture* the molten glass was first blown and spun into a long cylinder. To produce a flat sheet of glass the cylinder was cut open longitudinally and flattened out on a plane and hot surface.[33] However, with the exterior circumference of each cylinder being larger than the interior, the inner surface is forced to expand and the outer surface to contract during the flattening process, producing the characteristic uneven and creased appearance of sheet glass, describe as 'cockles' by glass manufacturers. [34] The power of the 'burning lenses' formed inside the glass varied significantly, due to great differences in quality of the glass.[35] A gardener, who had conducted detailed observation, reported that largest quantities of these lenses were not strong enough to cause any material injury to plants, but in some cases were extremely powerful and produced a 'sensible degree of heat on the hand when placed within foci'[36]

Other horticulturists, including Joseph Paxton and John Lindley, believed that insufficient ventilation was the chief cause of the scorching of plants. Due to lack of ventilation, they argued, plants were exposed to the 'dry and scorching heated air,' which was trapped inside the house. The *Gardener's Chronicle* reported that the use of sheet glass had significantly increased the issue of overheating in glasshouses as a result of the high quantities of solar heat that were admitted due to the increased transparency and the greater air-tightness of the glass envelope.[37] John Lindley wrote that laps in the glazing of crown glass houses were large and numerous, providing abundance of apertures through which the heat could freely escape. This, however, was no longer the case with sheet glass since the use of larger but fewer panes had significantly reduced the number of laps.[38] Incidents of overheating were reported to be more frequent and ventilation had to be provided much earlier to check the temperature of the internal atmosphere since it increased much more rapidly with the new glazing.[39] As a consequence, contemporary sources show, the careful control of sunlight and excessive solar heat gains became a more important concern in glasshouse design and lead to experiments with new ventilation regimes, shading systems and different types of glass that functioned as light filters or diffusers. The *Gardener's Chronicle* and *Book of the Garden* report that while some gardener's covered the lenses in individual panes of glass, others rendered the glass with different types of translucent varnishes in order to temper the excessive sunlight and heat during the summer period that could be washed off during the winter period, when sunlight was limited.[40] John Smith wrote that in the old glasshouses at Kew, shades, in the shape of mats or roller blinds, were only employed occasionally during extremely sunny summer days, mainly to protect specimens of particularly light sensitive tropical plants. With the introduction of sheet glass, however, the use of blinds became absolutely necessary in tropical houses.[41] William Jackson Hooker, however, rejected the use of blinds in the Palm House. Reporting that the use of canvas had already proven very 'troublesome and expensive' in glasshouse of small scale, Hooker wrote the Commissioners that blinds were completely impractical for the Palm House, due the large scale and complex geometry of the structure.[42] To avoid the use of shades, Hooker wrote the commissioners that tinted glass, as a light filter that excluded all the sunlight except the solar radiation responsible for injuring the plants should be use alternatively.[43] Contemporary sources show that the management of operable shades was seen widely as a very difficult operation.[44] In an attempt to avoid the use shades, the *Gardener's Chronicle* reported, experiments were conducted with rolled plate glass, which, due to its irregular ribbed surface diffused all of the direct sunlight before entering the glasshouse.[45] The development of tinted glazing for the Palm House, which will be discussed in the following sections, can be understood as part of this effort to appropriate the light conditions inside new glasshouses for horticultural purposes. The physicist Robert Hunt criticized that sheet glass exposed plants inside glasshouses to very unnatural light conditions, but argued that tinted glass could be used to reproduce the lighting that are more close those experienced by plants in their natural habitat.[46]

Lindley's call for a Scientific Inquiry

Alexander Milne of the Department of Woods and Forests, feeling unqualified to make any judgment about Turner's proposal, referred the matter to Hooker, asking him for his opinion about the practicality of the use of coloured glass.[47] In his reply Hooker wrote that he agreed with Turner, but advised the commissioners to seek

ISBN 978-0-9566139-1-2

the assistance of a scientist to specify a type of tinted glass that was suitable for the Palm House.[48] He wrote that a scientific study was necessary to ensure that the tinted glass would not only successfully protected the plants from excessive levels of solar radiation, but also that the composition of the filtered light admitted into the Palm House was not in any other way detrimental to the vigour of the tropical plants.[49] The importance of a scientific study had been raised by the botanist John Lindley with reference to recent research into the effect of filtered light on plant growth, which had shown that much caution had to be taken in the use of coloured glass in glasshouses.[50] The effect of light admitted through coloured glasses, research had shown, could be severe, even fatal, to the plants grown under it.[51] Hooker reported to the Commissioners that Lindley had recommended the appointment of the experimental physicist Robert Hunt as a scientific consultant, arguing that he was the only person in the country who possessed the required scientific expertise to conduct the study. Hunt's research was highly relevant to the specification of coloured glass since he had not only conducted extensive research on the composition and the various properties of light, but had also studied how the growth of plants was affected by light transmitted through coloured glass and coloured fluids held in glass cells. Detailed accounts of these experiments were published, in the *Researches on Light and its chemical relations* and paper presented to the Society of Arts and the British Association.[52] These accounts reveal that Hunt had already anticipated that special types of coloured glass could be used to purposively control the composition of sunlight transmitted into glasshouses in order to control plant growth. He had observed that growth was not only influenced by visible light, but also by two classes of non-visual radiation, which he described as *calorific*, which is solar heat, and *actinic* radiation, which was radiation that stimulated photochemical reactions.[53] He also discovered that the vigour of plants at the different stages of growth, such as germination, flowering and fruiting, depended on the influence of a particular class of sunlight.[54] The germination of seeds, for instance, was faster under the influence of isolated actinic rays than under the influence of the whole light spectrum.[55] Hunt also discovered that different coloured glasses, acting as light filters, could be used to regulate the proportion of light, heat, or actinic rays that were transmitted, and thereby became a means by which the light in which plants were cultivated could be carefully controlled.

Hunt's concept of 'selective glazing' was based on his detailed studies of the relative transparency of different coloured media to each class of radiation. His research had shown that the proportion to which each class of radiation was represented varied significantly with the colour of the light. Hunt produced various charts in which he mapped the respective intensity of thermal and actinic rays in different coloured light. Through the use of monochrome glasses, which only admitted light of a particular colour, it was therefore possible to also regulate the relative proportions of heat, luminous and actinic rays that were transmitted. It shows that the intensity of actinic radiation rays, for instance, was highest under the blue medium. Although it was not possible to completely isolate actinic radiation from light and heat, Hunt had demonstrated that the influence of luminous rays, which inhibited germination, was significantly reduced by means of blue glass in order not to interfere with the germination of seeds stimulated by actinic rays.[56] Hunt's commission to develop a special glazing for the Palm House at Kew, represented his first opportunity to actually apply his scientific theories and methods to the design of a horticultural building. The following two section illustrate that his research not only provided the scientific theories underlying the concept of 'selective glazing', but also the experimental methods that enabled Hunt to scientifically specify a glass of the required properties by simulating its actual performance.

Robert Hunt's First Scientific Report

Shortly after the receipt of Hooker's letter, the department of Woods and Forests approved Hooker's recommendations and commissioned Hunt to provide the commissioners with scientific advice on the feasibility of the use of tinted glass for improving the lighting inside the Palm House and to recommend a suitable type of glass.[57] On December 6th 1845, Hunt submitted a report with recommendations to the Commissioners entitled *Memorandum respecting the Coloured Glass most favourable for Glazing a Hot or Greenhouse*.[58] In this report he stressed that the first prerequisite to the selection of a glass for horticultural purposes was a detailed scientific understanding of the various constituents of sunlight and its respective influence on the physiology of plants. The second step was the identification of the specific light requirements of tropical plants and the third step the investigation into how these light conditions can be purposively created by means of tinted glass. Hunt, however, rejected the blue and violet glasses that Hooker and Turner had previously recommended.[59] While Hooker and Turner did not explain the reasons for their choices, Hunt reported that a Dr. Horner from Hull, had recommended violet glass as ideal for horticultural purposes, since it had been found to be not only effective in reducing the transmission of solar radiation but also to stimulate rapid growth.[60] Hunt, referring to the findings

ISBN 978-0-9566139-1-2

of his own experiments, rejected the use of blue and violet glasses, warning that it would interfere with the natural growth of plants inside the Palm House.[61] He reported that the *actinic* radiation, which is strongest in blue, indigo and violet light admitted by the proposed glasses, stimulated the germination of seeds and assisted the plants up to the development of the first leaves. However, in the growth stages beyond the period of germination, the influence of these rays interfered with the natural growth of plants. Hunt's warnings about the side effect of blue glass were informed by the findings of scientific experiments that he had conducted prior to his engagement at Kew. Detailed accounts of Hunt's experiments can be found in a paper read by Hunt at the British Association for the Advancement of Science in 1847.[62] Hunt writes that he had made extensive studies of the properties of the light passing through transparent glass trays filled with white, red, yellow and blue liquids and its respective influence on the growth of various types of plants that he had grown under these trays. The experiments revealed that blue glass, while admitting almost all of the actinic rays, not only intercepted heat rays but also large portions of visible light, but stressed that access to visible light was 'essential to enable the plants to effect the decomposition of the carbonic acid of the atmosphere and form woody fibre'[63] In two experiments he had demonstrated that plants that were grown under blue inhibited the decomposition of carbonic acid and thereby reduced the amount of woody fibre that was formed in stems and leaves. As the result the plants developed weak and extremely elongated stems. In his report Hunt concluded that conditions required for the healthful growth of plants could be achieved with a green glass of the correct tint. He suggested adding copper oxide into the clear sheet glass to give it green tint, which he describes as 'pea-green.' and claimed '*will admit light and chemical power in the same proportions as white glass, but it will obstruct the passage of those rays which produce the 'scorching' desired to be avoided.*'[64] It became a means to avoid the before mentioned side effects of blue and violet glasses, since green glass, in contrast to blue glass, did not obstruct much of the visible light that was vital to sustain the vigour of plants in the growth stages after germination.

Specification of a Special Palm House glass

Any further proceedings concerning the selection of coloured glass rested until June 1846, when Turner, intending to procure the glass from the manufacturer, inquired whether the Commissioners had come to any decision.[65] Although they had approved Hunt's recommendation in principle, the Commissioners insisted that careful research was required for the final specification of the glass. In July 1846 Hunt was asked to conduct, with the assistance of Burton and Turner, a detailed scientific study to specify more accurately the hue of green that was required to produce the desired lighting conditions.[66] The Commissioners wrote that 'adjustment of the tint may be of great importance, and involve some nicety in matters of which they consider you are particularly competent to judge.'[67] The caution from side of the commissioners has to be seen in the context of the fact that the recommendations in Hunt's first report were based entirely on the findings of past research. No experiments had hitherto been conducted specifically for the selection of the glazing and to test its performance. This shows that question of the glass was treated as a very serious matter by the commissioners. The project correspondence also show that the development of a special glasshouse glazing had only been possible through the unanimous support by the Commissioners of Woods and Forests and the provision of additional funding by the Treasury to cover the extra expenses incurred by the research and the manufacture of a special glazing.[68] The utilization of his scientific experiments for the specification of a special Palm House glass was the subject of Hunt's second report submitted on 11th January 1847.[69] Hunt, who had been supplied by Turner und Burton with thirty-seven samples of green glass from different manufacturers, submitted each sample to a set of three of experiments, which he had used in past scientific research, to measure its transparency to luminous influence, to ascertain the amount of obstruction caused to the passage of chemical rays and to measure the amount of solar radiation permeating each specimen. [70] While an optical prism, which enabled him to split the light into the different colours, was used to reveal the range of the visible light transmitted, the quantity of solar radiation transmitted by each sample was measured by means of a method formerly used by the scientist Sir John Herschel:

> *silver paper stretched on a frame was smoked on one side by holding it over a smoky flame, and then while the spectrum was thrown upon it the other surface was washed with strong sulphuric ether. By evaporation of the ether the points of calorific action were most readily obtained, as these dried out in well defined circles long before the other parts presented any appearance of dryness.*[71]

The result of this experiment, as illustrated in Figure 1, bar 3 and 4, was a visual image of the relative solar intensity across the entire spectrum of the transmitted light. The chemical change produced by the same spectrum, was measured by means of photographic preparations susceptible to chemical change under solar

ISBN 978-0-9566139-1-2

influence.[72] To study the degree of scorching, the discolouring of foliage caused by solar radiation transmitted by different samples of glass, Hunt had conducted an additional experiment. Instead of using living plants, as he did in his past experiments, Hunt used vessels filled with solutions of the green colouring matter abstracted from palm leaves[73] that had been supplied by John Smith on W.J. Hooker's request.[74] Referring to experiments by John Herschel, which had shown that the colour of leaves and flowers were destroyed by same class of rays that destroyed the green colouring matter of leaves, Hunt found that solutions were sufficient to simulate the relative scorching effect of the filtered light on actual living plants.[75] The relative scorching effect of the light transmitted by each glass sample was determined by measuring the length of time over which each solution lost its green colour under its influence.[76] However, according to Hunt's lecture at the Society of Arts in February 1848, the experiments had proven that none of these glass samples had fully satisfied the set performance criteria. Order to identify the most suitable tint Hunt conducted an additional series of experiments, in which he studied the composition of light passing through transparent fluids tinted with different chemicals. The use of coloured solutions enabled Hunt to control more subtly the shade and concentration of green tinting. At the end of these experiments he discovered that copper oxide, in a very diluted state, was the most suitable additive for the exclusion of heat.[77] In the specifications included in his final report, he instructed the glass manufacturer, apart from adding copper oxide in the right dose into the mixture of conventional glass, to remove the oxide of manganese used to whiten conventional sheet glass.[78] The exclusion of manganese was to prevent the glass from loosing its intended optical properties. In his report he stressed that manganese, under the sustained influence of sunlight, would cause the green glass to gradually whiten,[79] and at a lecture at the Society of Arts he warned that it would take on a pink tinge. The 'slightest approach to redness', he warned, 'would allow the free passage of those rays it was so important to obstruct."[80] To successful produce a glass with the desired properties several trials, conducted in collaboration between Hunt and the glass manufacturer Messrs Chance of Birmingham, were undertaken.[81] The specified glass, according to Hunt, was opaque to the non-visible infrared radiation and cut off the far end of the red in the visible spectrum, which caused the most severe destruction to the green colouring matter of leaves. Hunt reported that the colour of the solution had remained completely unchanged when exposed to light passing through a sample of the specified green glass but faded under the light transmitted through samples of white sheet glass.[82] At the same time, however, the experiments had verified that the specified glass had caused minimal obstruction to the actinic and visible light, required to ensure the vigour and natural growth of tropical plants.

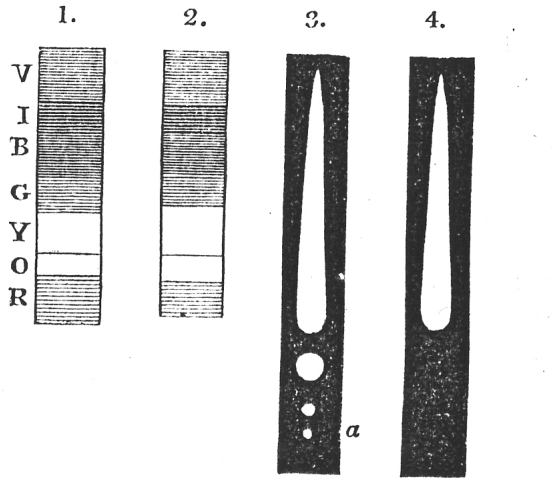

Figure 1: Graph showing the optical properties of the copper tinted glazing specified by Robert Hunt for the Palm House, showing the spectrum of visible light formed by a clear flint glass prism [1] and the same spectrum after it passed through a pane of copper tinted glass [2]. The green glass, Hunt reported, was found to cut a small portion of red and to reduce the length of the yellow spectrum by increasing the green. It also shows that the green glass [4] intercepted the intense infrared radiation transmitted by a clear flint glass prism, marked 'a' in graph 3. [Source: Hunt, Robert [1854] *Researches on light*, London: Longman, p. 383. / Cambridge University Library]

ISBN 978-0-9566139-1-2

The Office of Woods finally authorized the use of the tinted glass on 20th January 1847.[83] However, the project correspondence show that the final decision about the choice of glazing was not only based on scientific and horticultural criteria but that Burton was also requested by the Commissioners to evaluate Hunt's recommendation in terms of its impact on the architectural aesthetic of the Palm House.[84] Burton, who has been supplied with a sample, approved that the tinted glass would not have a negative effect on the appearance of the house, and Hunt stressed that the recommended glass would be more transparent than the more heavily tinted crown glass used in the old glasshouses at Kew, since it was '*so slightly tinted that it presents no appearance by either reflected or transmitted light materially*'.[85] A contemporary account in the *Gardener's Chronicle*, however suggest, that the green glass had a significant impact on the appearance of building. It wrote:

> *The glass employed there is of the most delicate emerald green. Viewed from within, and at right angles, it appears colourless, seen obliquely from within the green is apparent; viewed externally the tint is very decided, and when the sun strikes it at a particular angle it presents a most curious and beautiful appearance, the whole surface that is illuminated glowing like a mass of gigantic fiery opals*[86]

CONCLUSION: A History of Performance and Adaptation

This paper has shown that Hooker, the Commissioners, the architect and engineers relied on Hunt's scientific knowledge and experiments to test the optical properties as well as to determine the potential side effects of the light transmitted by the specified glass on the growth of vigour of plants. The correspondence also indicates that they had taken a great amount of trust in scientific experiments as means to make predictions about the actual performance of the glass. Since one of the primary objectives behind the Palm House design was to provide adequate lighting for nurturing tropical plants, the performance of the glazing was an important a criteria by which the success of the building would be finally judged. However, the use of green glass and the underlying scientific theories and methods also evoked some scepticism among contemporary critics. Smith in his *History of the Royal Gardens Kew* questioned whether the effect of green glass on vigour of actual Palm trees could be effectively simulated by the type of laboratory experiments Hunt had conducted.[87] In March 1848 the *Gardener's Chronicle* wrote that the question whether Hunt's theory was true or not would be answered in the future by the performance of the building itself.[88] The building's post-occupancy history was recorded in great detail in the *Kew Bulletin* and various contemporary scientific journals throughout the nineteenth and twentieth centuries. This history, which cannot be discussed here in much detail, reveals that the Kew authorities continuously examined the Palm House in terms its facility to provide adequate lighting, temperatures, humidity and ventilation for the cultivation of tropical plants.[89] It illustrates the difficulties connected with its dependence on natural light for supplying the plants with adequate levels of light and reveals that a condition of its success in terms of providing adequate lighting for the cultivation of a variety of plants, including plants that are sensitive to sunlight as well as those requiring large quantities of sunlight, was the facility of the building to adapt to a very changeable ambient environment. The latter proved very problematic and forced the authorities at Kew to finally abandon the use of tinted glass. According to Hunt's own account, no scorching had been observed inside the Palm House during the first five years of the building's use and the green sheet glass was also adopted in numerous other glasshouses at Kew, such as special houses for tropical ferns and orchids.[90] John Smith, while noting that the green glass had been very successful in providing good lighting for the plants, stressed that it did not entirely eliminate the need for shading. He reported that auxiliary shading was required to protect plants that are particular sensitive to sunlight, such as tropical orchids and ferns.[91] However after forty years Kew gradually replaced the tinted glazing with clear glass and the objective of the authorities at Kew was no longer, as in the 1840's, to moderate the transmission of sunlight but to exclude as little of the available sunlight as possible. But what did cause this change in strategy?

Contemporary sources show that the management became increasingly anxious about the sunlight being obscured by occasional fogs as well as by the general pollution of the atmosphere with particles of coal smoke, emitted by the city that was extending South-west towards Kew.[92] According to an article entitled *The Reduction of Town Fogs* the deficiency of sunlight in towns was considerable while the countryside, uninflected by urban pollution, was 'enjoying bright calm frosty weather.'[93] As a result of these changes the quantity of sunlight received by the plants inside the Palm House in the 1880's had significantly declined relative to the 1840's levels that had formed the basis for the original design.[94] Insufficient sunlight was reported to inhibit the growth and

ISBN 978-0-9566139-1-2

flowering of plants inside the Palm House and other glasshouses, in particular during the winter period.[95] According to the *Kew Bulletin*, it was after the extreme obscurity experienced during the winter 1885-86, that the authorities at Kew finally concluded that it was necessary in the future to maximizing the admission of as much of the available sunlight as possible and decided to stop the employment of green glass in all glasshouses, except from the tropical fern-houses and the Palm House.[96] The Palm House itself was not re-glazed with clear glass until 1930's when the issue of insufficient light had become even more severe.[97] In addition to the tint of the glass, deposits of black soot on the glass rendered the Palm House glazing extremely opaque. The *Kew Bulletin* wrote that while the 'sticky layers of soot' that settled on the glass during fogs were regularly washed off, black deposits, which had encrusted the upper part of the roof for several years, caused serious trouble.[98] The *Kew Bulletin* wrote that tinted glass was no longer needed since the combination of atmospheric pollution and glazing meant that light levels were too low for the cultivation of plants and alternative means of glazing and regulating the transmission of light had to be adopted. Kew went to great length to adapt the glasshouses, even the whole process of plant cultivation, to the changing environment. Various reports published in newspapers and Journals between 1886 and 1938 show that the authorities at Kew went to great length to study the influence of the declining sunlight levels and took measures to increase the lighting inside glasshouses, involving experiments with boosting the light levels by means of clear glass and electric lighting.[99] In this context Kew also commissioned a re-examination of the optical properties of the type of tinted glass that was currently used in various glasshouses. However, the glass was not identical with the glass originally specified by Hunt and was found to intercept too much sunlight.[100] Unfortunately there appears to exist no evidence of the performance of Hunt's original glass. The *Kew Bulletin* reported that while the Palm House was originally glazed with green glass especially manufactured in accordance with Hunt's specifications in 1847 the problem in the subsequent forty years was that not enough caution had been taken in obtaining a glass that adhered to the standard of the original glass. The green glass that was procured for repairing the original Palm House glazing in subsequent years had a much darker tint. A scientific study conducted by the astronomer Norman Lockyer in the 1880's, using new methods for measuring light, had revealed that the new green glass, in contrast to Hunt's original glass, severely reduced the intensity of sunlight across the whole spectrum.[101] Measurements of the light transmitted through this glass had shown that red-yellow and blue light was reduced by nine tenth in intensity, while the light with the highest intensity was yellow-green.[102] The *Kew Bulletin*, argued 'that the green glass in recent use at Kew, according to modern accepted data of vegetable physiology intercepted about half the effective influence of ordinary sunlight on the processes of plant life' The author of this report also questioned the validity of Hunt's original research, referring to more recent experiments by the French scientist Zacharewicz, who studied the effect of clear and coloured glass on the growth of Strawberries and had discovered that 'ordinary clear glass gave the best and earliest fruit'[103] These studies, the author claimed, demonstrated that green, red and blue glass interfered with the healthful developments of plants. He stressed that this showed that there was no evidence that green glass, as Hunt had originally claimed, provided better light for plant growth than blue glass.[104]
Furthermore the *Kew Bulletin* explained that it was found almost impossible to procure a glass that, as Hunt had originally specified, was absolutely free from manganese, and the green glass that was later used at Kew was also tinted with Iron-protoxide rather than with copper oxide.[105]According to the architect Maurice B. Adams the tinted glass at Kew, which had not been carefully manufactured, had lost its green colour and had taken on a pinkish tint.[106] The chemist William Crookes writes that the green colour of the glass was not permanent but, if exposed to sunlight for a long period, gradually lost its green colour.[107] It was reported that every stage from green through white to pink could be observed in the glazing of the Palm House, according to the respective age of each pane.[108] The impermanence of the tint, however, meant that the sunlight transmitted could longer be effectively controlled. The *Kew Bulletin* and *Gardener's Chronicle* wrote that Kew conduced experiments with clear glass in a number of glasshouses, which had demonstrated that even very light sensitive plants, such as tropical ferns, could be successfully cultivated under clear glass with a system shades, and from the 1890's it became determined to abandoned the use of tinted glass altogether.[109] While plants requiring the maximum amount of sunlight could be supplied with more sunlight, operable shades were still required to provide, at times, protection for light sensitive plants. This showed that the use of tinted glass as a lighting control strategy had proven very inflexible, offering no means of increasing the transparency to sunlight during the winter, when sunlight was only available for a few hours and was low in intensity, neither did it take into account the different requirements of plants living naturally in the shade and those growing under fully exposure to the tropical sun, which Charles McIntosh highlights as an important consideration in the design tropical houses.[110] Various contemporary accounts show that the vigour and growth of plants inside the Palm house and other glasshouses was also very

ISBN 978-0-9566139-1-2

dependent on ambient lighting conditions. These varied significantly between the seasons and between the years.[111] In contrast to the use clear glass with blinds, which can be adjusted to changing lighting conditions, the use of tinted glass represented a very static approach to lighting control. In its conception the need for adaptation to a very changeable ambient environment, let alone the impact of atmospheric pollution, was not anticipated.

REFERENCES

[1] The design theories and principles are discussed in detail in *Magazine of Botany*, *Gardener's Chronicle*, *Transactions of the Horticultural Society of London*, *Journal of Gardening,* Loudon, John C., *Remarks on the Construction of Hothouses*, (London: J. Taylor, 1817) & Loudon, John C., *Encyclopaedia of Gardening*, (London: Longman, Hurst, Rees, Orme, and Brown, 1835)

[2] Ray Desmond, 'Who designed the Palm House at Kew Gardens?,' *Kew Bulletin*, 27, 2 (1972) 287-93 & Peter Ferriday 'Palm House at Kew, *Architectural Review*, February (1957) 127-8 & Diestelkamp, Edward J. The Design of the Building of the Palm House, Royal Botanical Gardens, Kew, *Journal of Garden History*, 2, No. 3 (July Sept.1982) pp. 232-272. & John Hix, 'Richard Turner: Glass Master,' *Architectural Review*, 152, 909 (1972) 287-93.

[3] Original correspondence is held at the archives of the Royal Botanic Gardens Kew and the National Archives. Correspondence with the scientist Robert Hunt are preserved in Hunt, Robert, *Researches on light in its chemical relations*, (London: Longman, Brown, Green, and Longmans, 1854) Appendix IV

[4] Note: List of parties involved in the design process:
Client: Commission of Woods, Forest and Land Revenues:
Garden Directors: William Hooker : Botanist and Director of the Gardens from 1841-65,
Designers: Decimus Burton: Architect/ project manager employed by the Office of Woods & Forests
Richard Turner: Engineer-Iron-contractor: Hammersmith Ironwork in Dublin
William Andrews Nesfield: Landscape Architect
Manufacturers:Messrs Chance & Co., glass-manufacturer, Birmingham
Easton & Amos, Malins, London, wrought-iron manufacturer
Consultants:
Turner´s Foreman: structural engineering and manufacturing
Prof. Robert Hunt: experimental physicist
John Lindley: botanist
John Smith: Experience practical gardener, Curator of Kew Gardens from 1841-1864.

[5] On the Agency of Solar light, *Magazine of Botany*, 5 (1839), pp. 203-206 & 'Of Gardening as a Science, *Magazine of Botany*, 6 (1840), pp. 33-37, pp. 59-63, pp. 155-59.

[6] Hunt, Robert, 'Memorandum of the Principles upon which the Glass has been selected for the Palm House in the Royal Botanic Gardens at Kew,' 11th January 1847 in Hunt, 1854, Appendix IV & 'Society of Arts', *Gardener's Chronicle*, 9 (1848), p.139 & Hunt, Robert 'Report on the Influence of Solar rays on the growth of Plants,' in *Report of British Association* (London: Richard and Taylor, 1847)

[7] Note: Richard Turner on the glazing to be used for the Palm House, according to John Smith, 'Green Glass, *Gardener's Chronicle*, 323, 8 (1880), 307-8.

[8] Letter from R. Turner to A. Milnes, 15 November 1845, in Hunt, 1854, Appendix IV.

[9] Ibid.

[10] Smith John, *History of the Royal Gardens Kew*, (unpublished Manuscript, Royal Botanic Gardens Kew, 1880) folio 61f.

[11] John Smith 'Green Glass, *Gardener's Chronicle*, 323, 8, (1880) 307-8

[12] Smith John, *History of the Royal Gardens Kew*, folio 61f.

[13] Diestelkamp, Edward J., 'Richard Turner and the Palm House at Kew Gardens', in Structural Iron, 1750-1850, Studies in the History of Civil Engineering, ed. by R.J.M. Sutherland (Aldershot: Ashgate, 1997) & Minter, Sue, T*he Greatest Glass House*, *The Rainforest Recreated* (HMSO, London, 1990)

[14] Report by D. Burton to W. J. Hooker, 7-22 March 1844 [National Archives: WORK 16/29/8: Palm House: Erection]

[15] Ibid

[16] Letters from R. Turner to W. J. Hooker, 25 May 1844, 23 November 1844 & Letter from D. Burton to W. J. Hooker, December 1844. [Director's Correspondence, Vol. 22, 1844: K-Z & Vol. 21, 1844 A-J]

[17] These heavy ribs are shown in Burton's original working drawings of 1844 [National Archives: Contract Roll, Palm House Kew, dated 27. 8. 1844, Work 13/III/ Kew Palm House, Grissel and Peto]

[18] Smith, John, *History of the Royal Gardens Kew*, folio 61f. & Smith, John 'Green Glass,' *Gardener's Chronicle*, 323, 8 (1880) pp.307-8.

[19]Guthrie, J. L., Allen, A & Jones, C.R. Royal Botanic Gardens, Kew: Restoration of Palm House, in *Proceedings of the Institute of Civil Engineers*, 1988, 84, pp. 1145-91

ISBN 978-0-9566139-1-2

[20] 'The New Palm House, Kew Gardens,' *Lady's Newspaper*, January 22, 1848 & Hix, John, *The Glass House* (London: Phaidon, 1996), p. 148.
[21] Smith, John 'Green Glass,' *Gardener's Chronicle*, 323, 8 (1880) pp.307-8.
[22] Letter from J. Lindley to W. J. Hooker of 6th April 1844 [Botanic Gardens Kew: Director's Correspondance Vol. 22, 1844: K-Z]
[23] Letter from R. Hunt to Commission of Woods and Forests of 16 December 1846 [WORK 16/29/8: Palm House: Erection]
[24] Conversations concerning the form was recorded in the following letters: Letter from Turner to Hooker of 28th February 1844, Letter from R. Turner to W. J. Hooker of 27th February 1844 [Director's Correspondence Vol. 22, 1844: K-Z], Report by D. Burton to W. J. Hooker, 7-22 March 1844 [WORK 16/29/8: Palm House: Erection]
[25] Loudon, John C., *Remarks on the Construction of Hothouses*, (London: J. Taylor, 1817) George Mackenzie: Paper: "On the form which the glass of a forcing house ought to have, in order to receive the greatest possible quantities of rays from the Sun', dated 1st August 1815.
[26] Loudon, John C., *Encyclopaedia of Gardening*, (London: Longman, Hurst, Rees, Orme, and Brown, 1835)
[27]This issue was discussed in Burton's report of 7th-22nd March 1844 and reference to the underlying theories can be found in: 'On the Agency of Solar light,' *Magazine of Botany*, vol. 5. 1839, pp. 203-206 & 'Of Gardening as a Science,' *Magazine of Botany*, vol. 6, 1840, pp. 33-37 & 59.63 & 155-59.
[28] Paxton's technical experiments are extensively discussed in the *Horticultural Register* and *Magazine of Botany* between 1835 and 1840.
[29] *Gardener's chronicle*, 33-1848 [12th August 1848] p. 539 & *Gardener's chronicle*, 34-1848 [19th August 1848] p. 555 & p. 559.
[30] *Gardener's chronicle*, 33-1848 [12th August 1848] p. 539 *&* *Gardener's chronicle*, 34-1848 [19th August 1848] p. 555 & Beta ,On Glass Burning,' *Gardener's Chronicle*, 17th March 1849, p. 166
[31] Sheet glass, Gardener's chronicle, 34-1848 [19th August 1848] p. 559 & Scorching effects of British Sheet glass, *Gardener's chronicle*, 24-1848 [10th June 1848] p. 381
[32] 'Scorching effects of British Sheet glass,' *Gardener's chronicle*, 27-1848 [1st July 1848] p. 437.
[33] Marson, Percival, *Glass and Glass Manufacture* (London: Pitman & Sons, 1918) pp. 91-93.
[34] *Gardener's chronicle*, 33-1848 [12th August 1848] p. 539
[35] 'Scorching effects of British Sheet glass,' *Gardener's chronicle*, 27-1848 [1st July 1848] p. 437 & 'Scorching effects of British Sheet glass,' *Gardener's chronicle*, 31-1848 [29th July 1848] p. 511
[36] 'Scorching effects of British Sheet glass,' *Gardener's chronicle*, 24-1848 [10th June 1848] p. 381
[37] 'Scorching effects of British Sheet glass,' *Gardener's chronicle*, 23-1848 [3rd June 1848] p. 365
[38] *Gardener's Chronicle*, 10-1848 [4th March 1848], p. 155.
[39] 'Scorching effects of British Sheet glass,' *Gardener's chronicle*, 23-1848 [3rd June 1848] p. 365
[40]'Varnishing Sheet Glass to prevent burning', *Gardener's Chronicle*, 14th April 1849, p. 230 & Glass Burning, *Gardener's chronicle*, 37-1848 [9th September 1848] p. 605. & 'Sheet glass,' *Gardener's chronicle*, 34-1848 [19th August 1848] p. 559
[41] John Smith 'Green Glass, *Gardener's Chronicle*, 323, 8, [6 March 1880] pp.307f. & Letter by W, Hooker to Commission of Woods and Forests of 8th January 1847
[42] Letter by W. J. Hooker to Commission of Woods and Forests of 8th January 1847 [WORK 16/29/8: Palm House: Erection]
[43] Letter from W. J. Hooker to Woods and Forests, 23rd November 1845, in Hunt, 1854, Appendix IV.
[44] Macintosh, Charles, *The Book of the garden*, Vol I (Edinburgh : William Blackwood, 1853), p. 295.
[45] *Gardener's chronicle*, 33-1848 [12th August 1848] p. 539 & *Gardener's chronicle*, 34-1848 [19th August 1848] p. 555
[46] Letter from R. Hunt to Commissioners of Woods and Forests, 16 December 1846. [WORK 16/29/8: Palm House: Erection]
[47] Letter from A. Milne to W. J. Hooker, 19th November 1845 [WORK 16/29/8: Palm House: Erection]
[48] Letter from W. J. Hooker to Commissioners of Woods and Forests, 23rd November 1845, in Hunt, Robert [1854] Researches on Light, p. 376f.
[49] Letter from W. J. Hooker to Woods and Forests, 23rd November 1845
[50] Ibid
[51] Ibid
[52] Hunt, Robert, *Researches on light in its chemical relations*, (London: Longman, Brown, Green, and Longmans, 1854) Appendix IV
& Society of Arts, *Gardener's Chronicle*, 26th February 1848, p. 139 &
[53]Ibid
[54] Robert Hunt was Professor of experimental physics at the School of Mines in Jermyn Street, London
[55] Hunt, Robert, *Researches on light in its chemical relations*, (London: Longman, Brown, Green, and Longmans, 1854) p. 225.
[56] Ibid
[57] Smith, John, Green Glass *Gardeners´ Chronicle*, 13, 323 (1880), p. 307-08 & Letter from T. S. Phillips of the department of Woods and Forests, to Henry de la Beche, 6th December 1845, in Hunt, 1854, Appendix IV.

ISBN 978-0-9566139-1-2

[58] Robert Hunt, 'Memorandum respecting the Coloured Glass most favourable for Glazing a Hot or Greenhouse,' 6th December 1845, in Hunt, 1854, Appendix IV.
[59] Letter from W. J. Hooker to Commissioners of Woods and Forests, 23th November 1845 & R. Turner to A. Milnes, 15th November, 1845. [National Archives: WORK 16/29/8: Palm House: Erection]
[59] Macintosh, Charles, *The Book of the Garden*, Vol 1 (Edinburgh: William Blackwood, 1853) p. 535. & Hunt, Robert, *Researches on light in its chemical relations*, (London: Longman, Brown, Green, and Longmans, 1854) p. 377.
[60] Ibid
[61] Robert Hunt, Memorandum respecting the Coloured Glass most favourable for Glazing a Hot or Greenhouse, 6th December 1845
[62] Robert Hunt (1848) Report on the Influence of Solar rays on the growth of Plants, British Association, Richard and Taylor.
[63] Ibid, p.25
[64] Ibid & Hunt, Robert 'Memorandum respecting the Coloured Glass most favourable for Glazing a Hot or Greenhouse,' 6th December 1845
[65] Letter from R. Turner to D. Burton of 29th June 1846 & Letter from D. Burton to Commissioners of Woods and Forests of 1st July 1846 [Archives Royal Botanic Gardens, Kew]
[66] Letter by R. Hunt to A. Milne of 10th July 1846 & Letter by Milne to Hunt of 9th July 1846, in Hunt, 1854, Appendix IV.
[67] Letter by A. Milne to R. Hunt of 9th July 1846.
[68] Letter from D. Burton to Commissioners of Wood and Forest of 27th March 1847[National Archives: WORK 16/29/8: Palm House: Erection] & B. Adams, Maurice, Art in the Conservatory and Greenhouse, 31 January 1880, p. 140 & New Palm House – Botanic Gardens Kew, Statement of the Amount provided by Votes of Parliament for the above Service and of the probable ultimate Charges against the same by Office of Woods of 7th April 1847 & Notes by D. Burton of 27 March 1847 [National Archives: WORK 16/29/8: Palm House: Erection]
[69]Robert Hunt, 'Memorandum of the Principles upon which the Glass has been selected for the Palm House in the Royal Botanic Garden Kew, 11th January 1847.
Civil Engineer and Architect´s Journal, 10, 34, pp. 287. & Letter from R. Hunt to Office of Woods, 16th December 1846.
Ibid.
[72] Hunt, Robert, 'On the Principle upon which the Tinted Glass used in the construction of the Royal Palm House at Kew has been selected, and the use of Coloured Media in the Cultivation of Plants**'** presented to the Society of Arts on 16 February 1848
[73] **Note:** Leaves from the following types of palms were used: Chamaerops humilis, corypha australis, latania borbonica, phoenix dactylifera [Smith John, History of the Royal Gardens Kew (unpublished Manuscript, Archives, Royal Botanic Gardens Kew, 1880) folio 62]
[74] Smith John,1880, folio 62 & Hunt, Robert [1854] p. 383
[75] Hunt Robert, 'Memorandum of the Principles upon which the Glass has been selected for the Palm House in the Royal Botanic Garden Kew, 11th January 1847.
[76] Letter by Hunt to Office of Woods of 16th December 1846
[77] Society of Arts, *Gardener's Chronicle*, 26th February 1848, p. 139 [9-1848]
[78] The Great Exhibition, *The Morning Chronicle*, March 23th 1851, p. 2.
[79] Letter by Hunt to Office of Woods of 16th December 1846
[80] Society of Arts, *Gardener's Chronicle*, 9-1848 [26th February 1848], p. 139
[81] Hunt, Robert. 'On the Coloured Glass employed in Glazing the new Palm House in the Royal Botanical Garden at Kew' *Civil Engineer and Architect´s Journal*, 10, 34, pp. 286f. (1847) & Society of Arts, *Gardener's Chronicle*, 26th February 1848, p. 139 [9-1848]
[82] Letter from Hunt to Office of Woods of 16th December 1846 [National Archives: WORK 16/29/8: Palm House: Erection]
[83] Letter from Burton to Office of Woods of 27th March 1947.
[84] Letter from Hunt to Commissioners of Woods and Forests, 16 December 1846, Letter from Burton to Commissioners of Woods and Forests, 6 December 1847
[85] Letter from Hunt to Woods 16th December 1846 & Letter from Burton to Office of Woods of 6th January 1847.
[86] 'Influence of Light on Vegetation,' *Glasgow Herald*, 20 March 1848.
[87] Smith John, 1880, folio 62f.
[88] *Gardener's Chronicle*, 10-1848 [4th March 1848], p. 155.
[89] A brief account of the difficulties with sustaining an adequate climate can be found in Minter. Sue, T*he Greatest Glass House*, *The Rainforest Recreated* (London: HMSO, 1990)
[90] Hunt, Robert, *Researches on light in its chemical relations*, (London: Longman, Brown, Green, and Longmans, 1854)
[91] John Smith 'Green Glass,' *Gardener's Chronicle*, 323, 8, [6 March 1880] pp.307f.
[92]'Green Glass in Plant House,' *Bulletin of Miscellaneous Information*, February 1895, pp. 43-45
[93] Rollo Russel, 'The Reduction of Town Fogs, Nineteenth Century and After,' *A Monthly Review*, 51, 299 (1902:Jan.) p.131-43

ISBN 978-0-9566139-1-2

[94] *Bulletin of Miscellaneous Information*, Feb. 1895, p. 44
[95] J.D Hooker [1881] Report on the Progress and Condition of the Royal Gardens Kew, During the Year 1880, HMSO, London, p. 6.
[96] 'Green Glass in Plant House,' *Bulletin of Miscellaneous Information*, February 1895, pp. 43-45
[97] 'Review of the Work of the Royal Botanic Gardens, Kew, during 1935', *Bulletin of Miscellaneous Information* (Royal Gardens, Kew), Vol. 1935, No. 10 (1935), pp. 599-651 & Review of the Work of the Royal Botanic Gardens, Kew, during 1936 *Bulletin of Miscellaneous Information* (Royal Gardens, Kew), Vol. 1936, No. 10 (1936), pp. 533-58 & 'Review of the Work of the Royal Botanic Gardens, Kew, during 1938,' *Bulletin of Miscellaneous Information* (Royal Gardens, Kew), Vol. 1938, No. 10 (1938), pp. 473-529
[98] ,Miscellaneous Notes,' *Bulletin of Miscellaneous Information*, Vol. 1921, No. 1 (1921), pp. 30-48 & 'Appendix: Review of the Work of the Royal Botanic Gardens, Kew, during 1935', *Bulletin of Miscellaneous Information*, Vol. 1935, No. 10 (1935), pp. 599-651 & 'Tropical Department. – Palm House' *Bulletin of Miscellaneous Information*, Vol. 1937, No. 10 (1937), pp. 533f. Scott, Dawn Christine (1984) The Palm House and Temperate House at the Royal Botanic Gardens, Kew: a history, bibliography and guide to sources of information 1841 to date. [B.A. Thesis, Polytechnic of North London School of Librarianship and Information Studies, 1984]
[99] 'Review of the Work of the Royal Botanic Gardens, Kew, during 1936,' *Bulletin of Miscellaneous Information*, 10-1935, p.535-53, 'Tropical Department - Palm House,' *Bulletin of Miscellaneous Information*, 10-1937 [1937] pp. 533f & Coal Smoke Abatement Society', The Times, 21 November 1901, p. 15 & Rollo Russel, 'The Reduction of Town Forgs, Nineteenth Century and After,' *A Monthly Review*, 51, 299 (1902:Jan.) p.131-43
[100]Maurice B. Adams 'Green Glass,' *Gardener's Chronicle*, 323, 8 [6 March 1880] pp. 307f.,'Green Glass in Plant House,' *Bulletin of Miscellaneous Information*, February 1895, pp. 43-45,
[101]'Green Glass in Plant House,' *Bulletin of Miscellaneous Information*, February 1895, pp. 43-45.
[102] Ibid.
[103] Coloured Glass and Strawberries, *Gardener's Chronicle*, 462, 18 [2 November 1895] pp.524f.
[104] 'Green Glass in Plant House,' *Bulletin of Miscellaneous Information*, February 1895, pp. 43-45,
[105] Maurice B. Adams 'Art in the Conservatory and Greenhouse', *Gardener's Chronicle,* 318, 8, [31 January 1880] pp.139-41. William Crookes 'On the Colouration of Glass by Natural Solar and other radiations,' *Proceedings of the Royal Society of London*, 74 (1904-5) pp. 524-28.
[106] Maurice B. Adams 'Art in the Conservatory and Greenhouse', *Gardener's Chronicle,* 318, 8, [31 January 1880] pp.139-41.
[107] William Crookes ,On the Colouration of Glass by Natural Solar and other radiations', *Proceedings of the Royal Society of London*, 74 (1904-5) pp. 524-28.
[108] Ibid
[109] 'Green Glass in Plant House,' *Bulletin of Miscellaneous Information*, February 1895, pp. 43-45, 'Tropical Fernery at Kew,' *Gardener's Chronicle*, 658, 26 [5 August 1899] p. 111
[110] Book of the Garden
[111]'Review of the Work of the Royal Botanic Gardens, Kew, during 1951', *Bulletin of Miscellaneous Information*, 1, 7 [1952] p. 18 & 'Review of the Work of the Royal Botanic Gardens, Kew, during 1948', *Bulletin of Miscellaneous Information*, 1, 4 [1849] pp. 1-18.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

BENEFICIAL OR HARMFUL: RETHINKING THE TIBETIAN YAK DUNG

Sihao SHEN

PhD Student, the University of Auckland, Auckland, New Zealand
s.shen@auckland.ac.nz

Keywords: Tibetan yak dung, Qinghai-Tibet Plateau, Australia cattle dun, Objective reality, Sustainable Environment

ABSTRACT

As a contribution to the living civilization of a branch of mankind, a unique cultural landscape, and a sustainable development concept, the Tibetan black yak hair tent featured a simple structure with fully handmade components. Most of its materials came from the yak, which also carried the tent from pasture to pasture. During an investigation trip to Tibet the question arose as to why cattle dung, so harmful to the Australian grassland, is a treasure in the Qinghai-Tibet Plateau?

This paper first looks at the natural setting of Tibetan vernacular civilization. The Qinghai-Tibet Plateau offers a hostile natural landscape for high altitude living. As part of this, the yak has been both indispensable and unrivalled in the Tibetan life. Nowadays even though some modern forms of energy have been introduced in Tibet, most Tibetan still burn yak and cattle dung as an essential part of daily life. The paper then compares attitudes to dung in Tibet and the Australian grasslands. Finally the primary theme of this paper is reached, which is that in any research analyzing objective reality or fact has to be done foremost.

In dealing with objective reality it is essential to suit any measures to differing conditions in terms of locality, time, and the people involved. This runs counter to much modern thinking and activity which is based on ideas of globalization and universality. The same issues can be found in architectural design. This means for sustainable design well researched environmental issues that pertain to the place and time and local culture will be the key to the success.

INTRODUCTION

> *Cow dung, cow dung, valuable cow dung, you are more precious than gold; you are burning yourself out, offering warm and bright to people.*
>
> *- Tibetan folk music 'Song of cow dung'*

Tibetan yak dung, as one of side products of yak economy and plateau ecology, is not only the fertilizer in production and fuel in the life, also the mascot in folk festivals and the treasure in Tibetan mind. The cow dung in Tibetan sentiment, status and role in above mentioned short song, has been appeared without doubt.

Coincidentally, yak dung in China's Tibetan areas is seen as treasure to be used rationally and efficiently, while on the other hand Australia once had serious fecal harm issue. Australia is a country of a myriad of luxuriant grasslands. From the 18th century, European immigrants introduced a large number of cattle and other livestock. The superior natural conditions supported the development of husbandry. But, because the cattle dung was not been treated effectively, a growing quantity of dung covered the developing grasslands causing a serious impediment to the growth of grass. So many blocks of bare patches appeared in the gradually degraded grassland.

ISBN 978-0-9566139-1-2

Being same as cattle's excrement, how can they present such striking contrast? What kind message the phenomenon is delivering and experience and lessons are worth of note and draw to academicals especially to some practice-oriented subject such as architecture?

QINGHAI-TIBETAN PLATEAU'S SKY, EARTH AND LIFE

A vast and sparsely populated area

Generally, the Qinghai-Tibetan Plateau with its very high overall elevation (an average elevation of over 4,500 metres) is known as the 'roof of the world' and 'the third pole of the earth', as some of the harshest geography on the planet is found there. It is a vast, elevated plateau in Central Asia covering most of the Tibet Autonomous Region (TAR) and Qinghai Province plus parts of Gansu, Sichuan and Yunnan Provinces in China. It occupies an area of 2.5 million square kilometres about four times the size of Texas or France. According the 1990 official census, there are 4,594,188 Tibetans living on the plateau, with 53% living at an altitude over 3500 metres. A large number people (about 600,000) live at an altitude exceeding the average 4500 metres in the Changtang-Qingnan area. Compared with the population density of 34.8 persons per square kilometre in Texas and 115 persons per square kilometre in France, Qinghai-Tibetan Plateau's population density is very low at only 3 persons per square kilometre . The Ngari (Pinyin: Ali) Prefecture, located in the western Qinghai-Tibet Plateau with an average population of 0.23 persons per square kilometre, is the region with the lowest population density in Tibet and China.

Although the Peruvian city of La Rinoconada at 5100 metres above sea level has been named by National Geographic in 2003 as the highest permanent human habitation it is not permanently occupied. The settlement of about 30,000 people located high in the Peruvian Andes is home to people who survive the lack of oxygen and high altitudes to work in the gold mine here. The town is situated at the base of a still operating gold mine, where almost all of the inhabitants work. Most residents, however, come and work for short periods of time before descending to lower, more hospitable elevations. It was reported by the China Tibet Information Center that the world's highest village, 5,070 metres above sea level, is located at Tui Village in Dalong Town, Lhoka Prefecture in the Tibet Autonomous Region of China. In Northern Tibet, the town of Nagqu, governs more than 40 million square kilometres of the vast territory at the average altitude of 4,507 metres of Nagqu Prefecture, and could be the place of the world's second highest cities in the heart of the Plateau. As a northern Tibet livestock and native products distribution logistics center and highway hub in the Tibetan pastoral areas, the town itself has an area of 0.4 square kilometres and jurisdiction over 28 villages (neighborhood) committees with a population of 17,000 in 1996, mostly Tibetans who have lived there for generations. There are 12 industrial enterprises, mainly ethnic handicrafts and raw materials processing industries. Considering many tourists complain about altitude sickness when visiting places like Denver, Colorado, located just over one mile or 1609 metres above sea level , it would seem nearly impossible for anyone to actually survive at an elevation of almost 4,507 metres. In fact, Nagqu town sits fairly close to the maximum altitude of human survivability. There are many arguments about the maximum altitude of human survivability all over the world, but one point is crucial, *"…, the main reason why communities do not live higher in countries like Nepal and Tibet is that there is little point in doing this. Crops and vegetables will not grow at these great altitudes, and it is not possible to graze livestock for 12 months of the year because of the lack of fodder. Therefore, there is no advantage in trying to develop a settlement there. By contrast, if there is another economic activity such as mining, it is indeed possible for a community to live at an altitude as high as 5100m.*" So the altitude of the highest permanent human habitation is probably determined by economic factors, rather than solely by human tolerance to hypoxia.

A harsh natural environment

The Northern Tibet plateau is different from any other grazing land in the world. Research has so far failed to convent this alpine land into arable farmland and orchards. In this non-farming land, the nomad could be the only solution to the need to subsist. With their permanent flocks of sheep and herds of cattle, the nomad becomes an element of nature, working in partnership with the sun, moon and snow, or as an adjunct of the natural world. A tent plus a group of cattle and sheep could be all that is needed for pastoral survival. Although the pastoral life year after year is not totally stable its pattern is reasonably predictable and hence it is a tranquil way of life. What interrupt the pattern are all kinds of natural disasters. In addition to floods and volcanic

ISBN 978-0-9566139-1-2

eruptions, Northern Tibet gets almost every evil: drought, insects, wind, hail, snow, debris flow, avalanche, earthquakes and fatal lightning strikes. Among these the snow fall in 1985 was very harmful. Between October 17th and October 19th heavy snow fell continuously for 41 hours in the Nagqu Prefecture. There was suddenly one metre thick snow north of the Tanggula Mountains. The sudden snowstorm caught the herders by surprise. It endangered life in150,000 square kilometers, and involved nearly 74,900 stricken herdsmen and 3,000,000 domestic animals in North Tibet.

Another super snow disaster in Tibetan history lasted eight months from September 1989 to April 1990, when continuous blizzards occurred in Baqen, Lhari, Biru, Bangor, Nagqu, Wenbu, and Sogxian. In this period 13 townships had their biggest snow disasters in history. Snow fell on over 120 occasions, and the entire pastures were covered to depth of between 50 and 100 cm. This affected 40,316 families, 222,300 people, and 5,097,500 livestock with a death toll of 980,000 animals. In some places the mortality rate was as high as 30%. Disasters caused direct economic losses of RMB 791 million. The snow covered pastures had sustained low temperatures below an average minus 30 degrees Celsius. Because the domestic animals did not have grass to eat for a long time, there was the phenomenon of live livestock eating dead livestock, big livestock eating small livestock and cattle and sheep eating tents in some hard-hit areas. The livestock mortality rate reached 50 - 70% at the partial township of Amdo and other hard-hit counties, and in some villages and households the mortality rate was an average 80--90%, meaning some places and households had no domestic animals left.

The pasture in northern Tibet may be the world's worst pasture. In particular, the western pastures at 5000 metres above sea level are semi-desert land. Fortunately, allegedly due to the strong sunlight and other reasons, the forage nutritive value is higher, which can be seen as compensation. However, the insurmountable hypoxia affects and stunts the growth of the livestock and, thus the western cattle and sheep are much smaller in size than the cattle and sheep of the low-lying eastern parts of Tibet.

The yak - the irreplaceable animal for Tibetans

Because nomadic settlements are scattered far apart and the natural high alpine and anoxic conditions are quite serious, the lives of the herders entirely depend on some kind of animal or vehicle for mobility. However, nature is fair and has created the most remarkable animal for the Tibetan environment, the yak. The yak is aptly referred to as the boat of the plateau or the camel of the snow and is also deeply involved with the living culture of the herdsmen. When people think of the animals of Tibet, the yak will probably be the first that comes to mind. There is a proverb in the Plateau which says that a Tibetan packs everything on the back of a yak. No one can deny that the yak is the lifeline of all the indigenous people of the highlands.

Besides being excellent pack and riding beasts, domesticated yaks are multi-purpose animals providing milk, fiber and meat. They can pull a plough, and even their dried dung is an important fuel. In addition, yak hair is woven into yarn and used to make tents and rope. Yak hide is used to make boots and boats. Yak meat is eaten by nearly every Tibetan family. It is high in protein with only one-sixth the fat of regular beef. In the summer months it is dried, but in winter it is often eaten raw. Yak milk is high in fat and is usually made into butter, yogurt and cheese. Even yak bones can be processed into bone powder and bone glue, the former for feeding animals or for fertilizer and the latter for industrial use.

YAK DUNG - AN IMPORTANT LINK IN THE ECOLOGY OF THE PLATEAU

Yak dung play an important role in Tibetan'

In the vast pastoral areas, the herdsmen move around in searching of pasture and living by the water and plant with black yak hair tent, whether winter or summer grazing pastures, leaving everywhere balls of yak dung. A small portion of them were saved by the herdsmen as fuel, most of them are scattered in the open country. After rain washed, the cow dung permeates into soil and causes the soil texture to be fertile, so soil is fertile and grass is lush. As a result, cattle and sheep can eat the fresh tender grass then be battened on the plentiful fodder. In agricultural areas, the ash from burnt cow dung is poured into the toilet and manure pits. After fermentation it becomes highly effective organic fertilizer, which discharges in the wheat and Tibetan barley fields, brings the agricultural harvest thus improves people's life. Cow dung is not preciousness what is it?

ISBN 978-0-9566139-1-2

Compared to the usage in production, the role of cow dung in life is more important. In Lhasa, Shigatse, Zetang and the surrounding areas where plains are scarce of trees, and there is no forest even in mountain, fuel is sorely lacking. So the general residents, farmers and herdsmen once mainly depended on cattle dung for cooking, cow dung was the most important fuel.

Nowadays even though some Tibetans get electricity from the grid or from solar collectors and there has also been a major switch to gas burners, most still burn yak and cattle dung for heating and cooking. The fact that yaks are much loved can be seen by their owners calling them "biogas converters". Generally about the size of a man's hand, yak dung paddies are laid on a mud wall for drying before they are used as fuel. About 7,000 paddies are needed by a household to get through the winter. Since yaks only eat wild grass from a variety of pastures, when the dung is burned it smells cleaner and sweeter than domestic cattle dung and burns very hot just like wood. Compared with the possibility of piles of dung patties polluting their drinking water, burning the patties is much better and less harmful. The Tibetan practice of recycling dung, not only effectively solves the fuel problem in a convenient manner, but also effectively solves the excrement problem which may be caused by leaving dung on the grassland.

More importantly, if the excrement was not used for the fuel, the few trees that grow would have to be cut for fuel. Tibetans have made great contributions to the protection of their ecological environment. This embodies the Tibetan spirit of love for nature and the principle of harmony between human being and nature.

In fact in most areas of the Plateau it is extremely difficult to grow trees and crops, and only wild prairie vegetation is widely distributed the high and cold landscape. Firewood is in extremely short supply and the sparse population and the inconvenient transportation methods mean other forms of energy are not available. The herders have long been using animal dung and grass as fuel in order to maintain a minimum standard of living. In the Tibetan language yak dung is called 'Qiuwa' which means fuel rather than excrement. Tibetan people do not think of the dung as unclean, they think it is very precious. In recent years yak dung has been selling in large quantities as a commodity, although this could significantly affect the ecosystem nutrient recycling.

Several years ago, a researcher thought that because of the struggle between man and grass for livestock excrement the lack of fertilizer was making the Tibetan pasture of lower quality, and the grassland was being degraded. Based on personal observation that the smell coming from burning yak dung is more or less like herbal medicine, it seems probable that there is not much N, P and K inside the dung. This conjecture gets validation from recent research:

> '*The total quantities of NO3--N, NH4+-N, available K and inorganic P from the initial dung largely decreased to almost none after 84 days. In the initial days (1~9 d), NO3--N, NH4+-N, available K and inorganic P in soil under the dung piles were found with a significant increase, and after two months the content of NO3--N was 8.4 times larger than that in the initial days. The contents of NH4+-N, available K and inorganic P at place 10 cm from the edge of dung piles increased significantly (P<0.05), but no significant increase was found (P>0.05) at 30 cm. Differing from the common belief, no obvious increases in organic C, total N and total P were observed in the whole experiment process.*'

AUSTRALIA CATTLE DUNG PRODUCED IN DIFFERENT ENVIRONMENT

Australia cattle dung was produced in a fully different environment. Australia has a latitudinal range of 11°S to 44°S, which , coupled with precipitation from 100 mm to 4 000 mm[1](), generates a wide range of grassland environments. The landscape is characterized by vast plains with only limited elevated areas. Native herbage remains the basis for a significant portion of the grazing industry. Most farms rely on animal products from grasslands. Arid and semi- arid tropical areas are used for extensive cattle grazing; water from artesian wells and bores is necessary. Pasture growth is very seasonal and stock lose weight in the dry season. The intermediate rainfall zone extends from south eastern Queensland through New South Wales, northern Victoria and southern South Australia, and includes part of south Western Australia. This high-rainfall zone forms the greater part of the coastal belt and adjacent tablelands of the three eastern mainland states; sown pastures are widely used. Sheep and cattle dominate livestock numbers; wool sheep and beef cattle predominate, but dairying is locally important.

ISBN 978-0-9566139-1-2

The Australian climate has following three essential features,

- The biggest proportion of arid area: the vast arid area is general and the proportion is the highest. The entire mainland average year precipitation is 470 millimeters which is the least in various continents except the Antarctica.
- A ring form distribution feature of the year precipitation: the Australian year precipitation's local distribution assumes half ring-like scheme opening the mouth to the west that means a reduction from the north, east, and south three sides towards interiors and western coasts. Rainy north has annual precipitation between 1,000 ~ 2,000 mm. This distribution pattern is unique among continents.
- Warmer and hotter weather covering the whole continent in general: Especially summer the interior area burning hot degree is not inferior to the Sahara area. Hottest month in January average temperature is above 30℃ in 3/10 areas of interior mid-west.

While the hottest place in the world is Dallol, Ethiopia, with the average temperature of 34.4□. [2]

The features of Australia topography are low, flat, and monotonous with weak cutting surface, gentle fluctuation and low plateau continent. Its average elevation of entire continent is 350 meters which is slightly higher than Europe. Entire continent of 96% areas are below elevation 600 metres (36% below elevation 200 meters). Because Australia has the unique terrain plus rich water source from underground lithosphere of Indian Ocean and Pacific Ocean and the vast quantity of artesian wells arose from high subsurface pressure. Therefore there are massive good forage grasses. Australia is the world's more developed countries of livestock by raising sheep and cattle primarily. The quantities of Australian existing sheep are 150,000,000 which occupy the world leader, and cattle herds are about 27,400,000. So, daily discharge of millions of cattle and sheep dung heap on the grassland.

The massive cow dung in the grasslands rapidly season and harden under the arid climate, and enable the pressing grass etiolating to die in large quantity. So that the lush prairie degenerated gradually and presented many bare spots. More seriously, it is also the culture medium for fly. The massive cow dung cause two kind of repugnant fly - buffalo flies and bush flies reproduce rapidly. Up into 1950s, flies were very rampant in the entire continent of Australia. Australia's prairie area was shrinking daily caused by serious pollution from cattle dung, Australia forage production was sharp decline. Because of the insufficiency of feed and spread of plant disease, livestock's yield and quality of meat, milk and fur were affected. Australia's animal husbandry has received the serious hindrance.

Although Australia has several hundred species of native dung beetles that make use of the fibrous dung produced by kangaroos, wallabies, wombats and other native mammals, they are unable to cope with the large quantities of dung produced by introduced livestock. So, the government allowed scientists to introduce more than 50 species of dung beetles to Australia from Europe and Africa to deal with the dung problem.

WHAT WE CAN LEARN FROM BOTH CASES

The tale and fate of both cattle dung are differ broadly which can be described by a Chinese idiom - 南桔北枳 (Nan jú běi zhǐ). The idiom means that orange growing in the south of Huaihe River is an orange while growing in the north of Huaihe River it becomes trifoliate orange. The both leaf are similar but their fruit's flavor are dissimilar. (Trifoliate orange 枳- zhi: deciduous shrub, the taste is painstakingly sour, sphere. Its hybrid with sweet orange is also known as Citrange.) The orange from south which transplanted into the north of Huaihe River will become trifoliate orange. Why is this so? Ii is because the water and soil for growing are dissimilar. The modern Chinese uses it to analogy the influence from environment to the human and objects, and describes some kind of change of thing and people along with the environment change and 'dissimilation' or

ISBN 978-0-9566139-1-2

'deterioration'. It is often used as metaphor of the same species, different environmental conditions because of the mutating of nature of things.

The fundamental enlightenment of this phenomenon gives us is that if people want to obtain expected effects for stable and sustainable development, at the foundation of well analysis of the internal conditions; especial analysis and research have to pay for external environmental conditions. Herein, the external environment is of the relationship between body and environment and the role of inter-ontology. The natural environment is the sum of all kinds of external environmental condition with related research body existence and development. The external environmental conditions are conditioned by many factors, in both dung cases, which could refer to climate, topography, precipitation, geographical location, natural resources, etc. The characteristics of the natural environment in general is not easy to control and becoming more and more volatile, so its influence is considerable, sometimes even to influence and dominate the trends and changes of research body. Such as, compared with the average elevation of 4,500 metres of Northern Tibetan Pasture, the Australian continent is 350 metres only. Another example is the Tibet Plateau belongs to the alpine cold meadow and steppe landscape, which is characterized by poor water and soil conservation and vulnerability to serious soil erosion for grass growing, while the Australian grasslands are characterized by savannah climate, but also the sub-tropical grassland climate. Arid and semi-arid tropical areas are used for extensive cattle grazing; water from artesian wells and bores is necessary. Pasture growth is very seasonal and evaporation rate is high.

In some cases the unintentional interactions of bodies with the adjacent main flow can cause the changes of external environmental conditions, which can be seen in 1970s' Australia rangelands. From 1950s to the early 1970s, this 20 remaining years were the big boom of Australia social economy development, namely in the historical third take-off in economic development. Precisely in this periods, the total numbers of Australian livestock increased dramatically from 1949-1950 cattle's 14.600,000 and sheep's 113,000,000 to 1979-1980 26.200,000, and 136,000,000 respectively[3] If the rate that the average cow drops 10-12 dung pads per day is calculated, fouling pastures and creating enormous breeding grounds for flies. A single dung pad can produce up to 3 000 flies within a fortnight[4]. Because of the accumulated livestock dung, the balances of pastoral ecology were broken which led the external environment from beneficent cycle to maleficent cycle. This case shows that there is a process from quantitative changes to qualitative changes for the correlation of bodies and environment. Any objects or conditions, through gradual accumulation of both time and space, could be changed by realization from quantitative changes to qualitative changes. So, a certain amount of things can keep their quality limits within certain extent. In a certain range the quality of things will not change with different amount, but once beyond this limit, the quality of things will change.

The case of Australia livestock dung is one of the best notes that quantitative changes increase to a certain degree will inevitably lead to a qualitative change. Well, what is the right aptitude to it. Another Chinese idiom防微杜渐(fáng wēi dù jiàn) ideally illustrate it. With the English equivalent of stitch in time saves nine, the original meaning of 防微杜渐 is that that arrest what seems to be the beginning of an unwholesome trend. This word was used in political related affairs at very beginning and now is widely used in almost everywhere, especially for medical issues. In Chinese traditional medicine as a preventive measure it reflects the principle of any ill-feeling has to be nipped in the bud before it develops into open revolt. This phrase tells us: hidden dangers to be removed in time to avoid a greater source of the disaster brewing health; diseases should be treated early to avoid greater harm to the body. Nowadays, under the influence of various factors in turn, Tibetan ecological problems have reached a line has to act decisively to arrest what seemed to be the beginning of a dangerous trend. The Plateau is threatened with various glacial recession, grassland desertification, and damage from an increasing rat population. If our alert lines of sight extend to the global, there are more environmental and ecological problems caused by the rapid urbanization and greenhouse effect. It is time to adopt a proactive and preventive approach, nip an evil in the bud with the aim at addressing problems that may affect a wide cross-section of our common ground. In the broadest sense of sustainability, the phrase "A stitch in time saves nine" still has significant meaning today.

ISBN 978-0-9566139-1-2

THREE PAIRS OF RELATIONSHIP IN DEALING WITH OBJECTIVE REALITY

Commonness and otherness

Although there is commonness of things and the environment, but it is relative; the otherness is actually absolute and is one of the basic features of things and the environment. On the one hand, it is impossible to find two natural condition identical things or environment; On the other hand, even tiny things and environment its internal differences still exist. Successful practice in one place, move to another country may be unsuccessful. Because of existence of otherness, different person are formed by different philosophy of the world or of human life which came from different life experiences, various vernacular culture came from different customs and aesthetic orientation. The later has been well demonstrated in the case of the Tibetans treated yak dung as precious as golden.

Investigation and research:

Most people consider they are equivalent and can replace each other; in fact, they belong to a process in two phases without particular order. Investigation is an activity for exploring or In-depth site visits to find the truth of objective reality, nature and law of development. The rigorous investigation can reveal the truth of body and environment and many facts emerged as a result of the investigation. To work out a right plan, one has to start with investigation. To make an investigation, one should go into matters deeply, not just scratch the surface. Research is a word that most people associate with science and objectivity. Skepticism, verification and even hypotheses can lead to investigation and research, and investigation is the beginning of research. Competent researchers follow no rigid scientific method in their investigations and made careful investigations of all problems that confronted him and never jumped at conclusions. Research is gathering of information and evidence which come from investigation. Researcher can only draw the conclusion by building on a thorough investigation.

Following the rules and breaking down the conventions

It seems that breaking down the conventions is an antithesis to following the rules , actually they can be coexisted. The law of the unity of opposites is the fundamental law of the universe. Learn the rules so you know how to break them properly. Observing the realistic and natural law, it is the basic principle for human progress and development. Anyone has violated realistic and the natural law will receive the penalty. If conventions cannot break down the conventions, innovation cannot be reached.

CONCLUSION

Though from different starting points, the two cases of livestock dung reached the same goal by different routes and drew almost the same conclusion. They offer full explanation that ecological rule is an objective law of nature. If people's economic activities in their own, not following ecological rules, only considering immediate successes and short-term gains, will destroy the ecological balance and even cause a vicious cycle in which people suffer disaster. However, once people understand and master the ecological laws, it should be beneficial to the ecological system, creating a virtuous circle of ecology and reaching a new ecological balance. Thus both economic and ecological benefits can be attained in a more reasonable, sustainable and prosperous circumstance.

REFERENCES

[1] McIvor, John G., Chapter 9, Australian Grasslands, Grassland of the world, J.M Suttie, S.G. Reynolds and C. Battello, Series title: FAO Plant Production and Protection Series - 34

[2] Osborn, L., World's Hottest and Coldest Places, © CurrentResults.com

[3] Notheal: 澳洲经济史之七：向发达国家转变时期的经济1950-1980□, 2009-07-26 09:29:08. Author: notheal, Seventh of Australian Economic History: economy of transformation time to a developed country (1950-1980) , 2009-07-26 09:29: 08, http://www.heybrain.com/notheal/article/3277.html

ISBN 978-0-9566139-1-2

[4] CSIRO Entomology - Black Mountain, *A problem caused by introduced livestock,* Exotic dung beetle solution to dung problem in Australia, http://www.csiro.au/solutions/DungBeetles.html

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

AN INTERDISCIPLINARY APPROACH TO FIRM CREDITWORTHINESS: A DEFENCE OF POLITICAL ECONOMY

D. Rajeev Sibal[1]

MPhil/PhD Candidate
London School of Economics and Political Science, London, United Kingdom
D.R.Sibal@lse.ac.uk

Keywords: Multinational Corporations (MNCs), State, Credit Ratings, State-Firm Relationship, Political Economy

ABSTRACT

This paper looks at the implications of a market phenomenon called piercing. Piercing is when a firm carries a credit rating higher than the state in which it is incorporated. It is argued that such firms represent the most successful firms in a domestic economy because of their low probability of default. What are the causal mechanisms by which a firm is able to achieve such levels of creditworthiness? Identification of the types of firms that are able to pierce is a challenge that existing theories have tried to explain but fail to do so accurately. This paper highlights the shortcomings of financial and economic approaches to piercing and suggests that a political economy approach should be considered instead. The importance of a political economy approach is reinforced through analysis of credit linkages between the state and the firm.

INTRODUCTION

> *Investors have been dumping the debt of struggling nations all over the globe, even as they show more confidence in many companies in these countries…The unusual result: It is now more expensive to buy protection for the debt of a range of countries than it is to purchase similar protection for some companies in those nations. It's as if investors are betting that the government of Spain, for example, has a greater chance of defaulting on its debt than do some Spanish banks.*
> Zuckerman, Wall Street Journal, 26 Jan 2010

The 'unusual' phenomenon whereby a corporation achieves higher levels of creditworthiness than the state in which it is incorporated has been receiving extra media attention during the recent financial crisis, but it is not a new conundrum for capital market analysts. In the world of emerging markets, this concept, called 'piercing', has existed for a number of years. Firm credit superiority is unexpected because states are generally considered to be more creditworthy than firms – governments control more assets, have greater revenue generating power, and will not disappear in the event of default. Nonetheless, piercing continues to sporadically occur.

Financial and economic theories have tried to reconcile this phenomenon by identifying piercing scenarios, but these theories fall short of accurately explaining what kinds of firms pierce and why. Financial explanations of piercing suggest that these firms are either able to generate more reliable cash flows than the state, or, from a political standpoint, have exceptionally strong ties to the state. Empirical analysis demonstrates, however, that the firms with the closest ties to the state are not necessarily the firms that pierce, and, in many situations, the cash flows of piercing firms could hardly be described as 'more reliable' than that of the state. Inconsistencies in

ISBN 978-0-9566139-1-2

financial explanations of piercing demonstrate the need for a more holistic approach. As a result, the paper will argue that a new approach is required, one that endogenizes politics and greater political economic structures.

Piercing firms are of interest because they represent the most successful firms in an economy. Creditworthiness is a judgment of an entity's probability of default, so higher creditworthiness is tied to a lower probability of default. Creditworthy entities have proven themselves as durable in the marketplace. How a firm is able to grow and achieve high levels of creditworthiness either in spite of or as a result of the domestic political economic environment is critical to deciphering economic development processes. Political economists have long argued that politics should play a central role in economic research, but an inability to empirically measure the extent to which politics inform business and financial markets has been a major source of criticism. Creditworthiness provides an empirical solution to measuring the impact of politics. Financial markets regularly assess the risk posed by both economic and political factors through the purchase and sale of debt securities. Credit ratings and credit spreads represent the perceived default risk associated with any given security. The paper uses firm and sovereign credit risk to link the role of politics and the domestic political economy to the market.

The state-firm relationship is central to piercing from both an empirical and theoretical standpoint. Empirical analysis clearly identifies sovereign risk as the principle driver of firm risk, even in instances of piercing.[2] Therefore, from a theoretical standpoint, how firms operate in the context of state policies is crucial to identify a more accurate account of the causality of piercing. Political economic theory should play a central role, since it directly addresses the state-firm relationship. Varieties of Capitalism (VoC), in particular, provides both a relevant theoretical framework and a methodological direction.

The remainder of the paper is organized as follows: section II defines core components of creditworthiness to demonstrate how the relationship between the firm and the state can be empirically identified; section III highlights the piercing paradox; section IV provides more detail about the firms that pierce; section V introduces the state-firm debate and identifies different explanations about how states and firms interact; section VI empirically demonstrates shortcomings of existing piercing explanations; and, finally, section VII concludes.

CREDITWORTHINESS AS AN EMPIRICAL LINK BETWEEN THE FIRM AND THE STATE

Creditworthiness is an empirical tool that allows the researcher to assess how factors such as democratic politics and political stability affect perceptions of risk.3 From the perspective of political science, it is well established that creditworthiness is as much a function of political risk as it is of economic and/or financial risk.4 Since the creditworthiness of firms and states is measured and priced on the same scale, this measurement becomes an important tool for analyzing the state-firm relationship. At both the firm and sovereign level, investors must assess endogenous and exogenous risk. Sovereign risk is a function of both the economic power of the state, which is determined by the aggregate performance of firms, the primary economic actors, and the political stability/capacity of the government.5 Firm risk is a function of cash flows, endogenously, and of the economic conditions of the state, exogenously.6 Before exploring more concretely the creditworthiness of specific firms and states, further detail about how credit is measured in financial markets should be considered. Two distinctions are equally important: the first relates to how creditworthiness is derived for states and firms and the second relates to how creditworthiness is measured in the marketplace.

The first distinction when dealing with creditworthiness deals with the differentiated ways in which state and firm credit risk is calculated. Sovereign risk has a strong qualitative component because of the political nature of government operations.[7] Quantitative approaches are certainly used in assessing sovereign risk, but political variables can easily undermine tendencies demonstrated by quantitative analysis.[8] Firms, on the other hand, tend to be judged first and foremost on their financial position, since financial analysis can approximate firm valuation and future cash flows.[9] Qualitative approaches are introduced through consideration of exogenous systemic risks. A notable amount of sovereign risk is also embedded in corporate risk, either directly or indirectly. Despite these quantitative and qualitative aspects, the market evaluates credit in a very uniform manner, either through the ordinal ratings of agencies or through credit spreads.

ISBN 978-0-9566139-1-2

The credit rating versus credit spread difference is a second consideration when working with creditworthiness. Ratings and spreads are the two ways that credit is measured in the market. Credit ratings are an assessment of credit risk as determined by a private agency. Agencies use a combination of quantitative and qualitative approaches to derive a rating, which is assigned based on a pre-defined ordinal scale.[10] Spreads, the alternative measure of risk, are the market pricing of risk. The price includes a risk free rate and a margin; the risk free rate is simply the cost of borrowing money with zero risk (frequently defined by US Treasuries), and the margin is the cost of the risk associated with lending to the firm or government based on the financial profile of said entity.[11] The credit rating and the credit spread,[12] in the long run, tend to be correlated.[13] Since the scales are uniform across entities, both provide a valid medium to compare firms and sovereigns.

THE PIERCING PARADOX

Leveraging creditworthiness to study the relationship between the firm and the state provides surprising results. In select cases around the world, corporations have better credit ratings or narrower credit spreads than the sovereign. The fact that firms are considered to have less risk of default than the state is a thoroughly unexpected reality. The reason that investors and credit rating agencies originally thought piercing to be impossible relates to how firms operate in the context of the macroeconomy, since the former are generally subject to the rules imposed by the latter. Interest rates, inflation, and currency policy, which are clearly in the policy domain of the state, directly impact firm liquidity and solvency. Firm dependency on the macroeconomy is especially relevant to, and most visible in, matters of liquidity. So, for instance, if the government were to default on foreign currency[14] debt because it did not have sufficient reserves, then it was assumed that firms in that economy would also have insufficient access to hard currencies (i.e. internationally liquid currency like the U.S. Dollar or Euro) to service foreign debt. At the most basic level, the assumption about state credit superiority is linked to the fact that the state wields great influence over the domestic economy through a number of different mechanisms.[15] The realities of states having greater tools, resources, revenue generating power and, most importantly, control over policies that direct the domestic economy, however, come into question as we more deeply examine the phenomenon of piercing.

Political scientists have not specifically engaged with the piercing phenomenon. Initial review might suggest that piercing is a boon to political theorists who argue that MNCs have grown to disproportionate sizes and can compete with states for power at an international level.[16] A natural question that arises is which firms have been able to pierce the sovereign and what is the nature of their relationship with the state? I hypothesize that empirical exploration of piercing firms will demonstrate relational ties between the state and the firm rather than the separation suggested by some existing political economic arguments. To be more precise, though, the nature of the state-firm relationship is indirect and not direct, as is suggested by financial and economic explanations of piercing. The firms that pierce are those that have best operated within the constraints of and responded to the comparative institutional advantages present in the domestic political economy.

WHICH FIRMS PIERCE?

The conditions for piercing to exist in an economy are a function of a range of macroeconomic and political factors. The principle focus here is not identifying those conditions that make piercing possible but rather to understand what firms in an economy pierce. By looking at a range of piercing firms, I can identify common characteristics. Another point that must be noted before taking the argument too far is that this argument is focused on domestically owned and operated firms. Piercing is too easily explained in a situation where a local firm has superior creditworthiness as a result of it being a subsidiary of a foreign MNC. The interest is in firms that have grown in the context of the local domestic political economy.

The leading credit rating agencies, S&P and Moody's, have permitted piercings since roughly 1997 and 2001, respectively.[17] The rating agencies were reacting to a market phenomenon where governments had defaulted yet domestically based firms had not. In the aftermath of emerging market crises in the late 1990s, it became apparent that the cash flows of certain corporations were outlasting that of the sovereign.[18] Certain firms had found ways to become more creditworthy than their home states.

ISBN 978-0-9566139-1-2

One example of realized piercing is in Uruguay, where in the midst of a financial crisis, the government differentiated which banks debts would be covered by government liquidity support. A combination of Argentine contagion and uncertain economic fundamentals led to a run on banks in the country and forced the government to intervene in the financial system. The financial crisis slowed economic growth, drained currency reserves, and forced the government to default on its debt.[19] The reason this is representative of a piercing scenario is because government support (and subsequent IMF intervention) was channelled to key domestic banks, while other banks were forced to fend for themselves. In such a crisis, where contagion was the key catalyst, one might expect the government to insulate itself from international markets by enforcing capital controls and taking over the entire financial system. The fact that Uruguay did not turn inward is representative of how integral the international market is for the Uruguayan economy. The fact that certain banks were allowed to operate while the government was in default demonstrates a real world justification of piercing.

Financial and economic theory soon began to reconcile piercing. Durbin & Ng present four pre-conditions for piercing to occur: firms must have international cash flows, access to a hard currency, foreign assets and/or guarantees, and/or political ties to the state.[20] The explanations of these factors are quite sound from a financial perspective. The first three reference a corporation's ability to access hard currencies, which insulates firms from sudden devaluations in the value of domestic currencies or assets. Firms are able to continue servicing debts without major interruptions in operations. Firms in the first three categories also tend to have a strong international market presence. The fourth pre-condition provided by Durbin & Ng is political, and, in many ways, it is a catch-all to explain situations not explained by financial and economic factors.[21] The theory posits that certain firms are sufficiently strategic to state interests that their operations are given priority over domestic finances. The political explanation, however, is limited in scope and is not tested.

Similar conclusions are made by Peter & Grandes in a study of piercing scenarios in South Africa.[22] In their single case country analysis, it is found that piercing firms are large, industrial multinationals and that firms in the financial sector are the most unlikely to pierce because of their strong ties to the state's sovereign risk (because of bank dependence on the macroeconomy, generally, and monetary policy, specifically). Political consideration is limited to the inclusion of sovereign risk as an independent variable and is not discussed as an explanatory consideration in the results. From the perspective of political science, the findings of Peter & Grandes are more limited than those of Durbin & Ng. What is compelling, however, is that even in piercing scenarios, sovereign risk is a key explanatory variable of the corporate risk premium: '*Sovereign risk appears to be the single most important determinant of corporate default premia*…'[23] Peter & Grandes expect this finding to extend to emerging markets, where sovereign risk remains a non-negligible risk consideration.[24]

The link between sovereign and firm risk is a subject that has recently been developed more substantially in econometrically based studies. Prior to the abovementioned studies, and not from a piercing perspective, Ferri, Liu & Majnoni[25] and Ferri & Liu[26] clearly identified the strong correlation between firm risk and sovereign risk in emerging markets. Similarly, Borensztein et al. find that sovereign credit ratings have a determining effect on firm credit ratings.[27] Borensztein et al. also expand and refine the principle findings of Durbin & Ng from a financial and econometric perspective but still fail to identify causality and recognize piercing patterns.

If sovereign risk is such a major component of corporate risk, in an emerging market context,[28] and sovereign risk assessment is an inherently political judgement, then the role of politics and the political economy should be central to our understanding of firm creditworthiness.[29] Indeed, Durban & Ng suggest that direct ties to the state or strategic value of a firm account for piercing scenarios unexplained by financial and economic theory.[30] But existing theoretical conceptualizations of politics in piercing scenarios are lacking and do not sufficiently explain empirical realities. A political economy approach to piercing will provide a more fruitful understanding of why it is that certain kinds of firms pierce.

THE FIRM AND THE STATE

Despite continued attention in both theory and practice, the nature of state-market-firm interaction remains mysterious. Many scholars agree that the state plays a role in firm development, but the specific ways in which that this occurs is not yet well researched and theorized. Management approaches reference state structures and organization as influences in firm development.[31] Economics takes a more systemic approach by trying to detail

ISBN 978-0-9566139-1-2

the market environments that are best suited for firm development. The Washington Consensus provided a classic example, with its advocacy of liberalization policies as a means to spur economic activity.[32] History, however, demonstrated that economic policy and capitalistic structures have to be adapted to individual state circumstances. 'Differential diagnosis' became a new catchphrase as economists reconciled failures of 'one size fits all' policies in the context of economic development.[33] As the Washington Consensus evolved into the post-Washington Consensus, core tenets of economic globalization began to be challenged. For example, a series of papers began to question the benefits of financial liberalization.[34] The inability for macro-policy to stimulate microeconomic development clearly demonstrates a weakness in the managerial and economic approaches to conceptualizing the state-market relationship.

From a political science perspective, many political economy studies do not venture very far beyond traditional economic policy prescriptions. Büthe & Milner add a political element to the study of FDI and reconfirm the economic viewpoint that engagement with the global economy helps developing countries attract international capital while raising prospects of economic growth.[35] Similarly, Simmons and Elkins reconfirm the increased pressures of economic competition in a global market.[36] One interesting finding of Simmons & Elkins, which has much broader implications than one might initially expect, is the fact that global norms do not seem to play a significant role in liberalization and policy diffusion. The lack of statistical evidence to support global norms as a driver of market structuring combined with the empirical reality of diverse forms of capitalism (that require uniquely structured economic policies) uncovers a principle weakness of many political economy theories that address economic policy planning, an inability to explain how capitalism manifests itself across states. One theoretical framework that has characterized differentiated forms of capitalism and engaged with the state-market interaction is the Varieties of Capitalism (VoC) approach because it focuses on the dynamism of the state-firm relationship.

VoC contends that since firms are the central actors in an economy, using the firm as a core analytical tool helps to explain institutional tendencies in the domestic political economy.[37] VoC borrows principles from management and microeconomics and applies these ideas in a political economic context. In essence, the macroeconomy is organizationally tied to the microeconomy through its institutions, which reflect efficiency trade-offs, and by firms that are responsive to macroeconomic incentives.[38] Recognition that firms operate in a market context[39] is a notion that tends to be overlooked in many political economy debates. Identifying operational ties between the macro- and microeconomy parallel the challenges that exist in tracing sovereign risk at the firm level. To trace credit ties between the firm and the state will be crucial to uncovering how firms pierce.

More specifically, VoC's analytical foundations provide an important theoretical frame of reference: firms are considered central actors in the economy and serve as a core analytical tool; the macro- and microeconomy are viewed as components of the same organization; institutional considerations are tied tightly to firm behaviour; and states are recognized as operating varied kinds of capitalism.[40] The implications of the above considerations for piercing are straightforward. First, the centrality of the firm parallels the firm focus of this paper, where piercing firms are the analytical centre. Second, the connections between the macro- and microeconomy are important to understand the relationship between the firm and the state, especially from a creditworthiness perspective. Macro-political and macroeconomic risks are two of the most important exogenous determinants of firm risk and behaviour. The third implication, which follows from the second, relates to the importance of institutions in constructing economic policies and regulating markets. Finally, the recognition that capitalism operates differently in different states is an important observation that helps to explain why different kinds of firms pierce in different states. Focus on the dynamic relationship between firms and states will facilitate clearer identification of causal mechanisms in piercing.

VoC-based studies identify typologies of capitalism that predict the kinds of firms that will be most successful in an economy by analyzing institutional complementarities and recognizing where domestic comparative advantages lie. The traditional example used to accentuate differences in capitalism is a comparison between the German (historically Rhineland) and American economies. In Germany, firms that specialize in incremental innovation tend to be the most successful because of the relationships that firms have established with labour, capital, and the state. The German economy is characterized as coordinated, or collaborative, which helps economic stakeholders collectivize risk. The United States, on the other hand, tends to prefer competitive arrangements and uses markets as a coordination mechanism. Firms compete for labour and capital, and the

ISBN 978-0-9566139-1-2

government maintains a high degree of separation from the marketplace. Because risk is not collectivized, firms find themselves under more pressure to perform, and, as a result, American firms tend to be better at producing products and services that are characterized by more radical innovation.[41] Similar distinctions are applied to a wide range of countries. For example, it is found that the Estonian and Slovakian economies are just as dichotomous, with Slovenia transitioning into a more coordinated economy and Estonia into a competitive one after the fall of communism.[42] The categorical typologies suggested by VoC are contentiously debated (and the extent to which VoC is useful beyond the OECD is still being researched), but the theory's ability to identify causal mechanisms of capitalist formation within each domestic political economy are compelling. The inter-disciplinary methodological foundations of VoC provide an avenue for more comprehensive understanding of capitalist varieties. VoC's focus on the firm and the relational constructions of economic stakeholders makes it especially relevant to study piercing firms.

AN EMPIRICAL EXAMPLE: (MIS) EXPECTATIONS OF PIERCING

An empirical example will help demonstrate the shortcomings of financial approaches to piercing and the need for a political economy approach. Brazil and India are examined because they have the highest number of piercing firms. See Table 1[43] for a comprehensive list of the piercing corporations.

Table 1: Piercing Firms in Brazil and India

Brazil Firm	Industry	BBB- Rating	India Firm	Industry	BBB- Rating
Aracruz Celulose S.A.	Natural Resources	BBB	Infosys Technologies Ltd.	Information	BBB
Banco Bradesco S.A.	Financial Institution	BBB	Reliance Industries Ltd.	Industrial	BBB
Banco Itaú S.A.	Financial Institution	BBB	Tata Consulting Services Ltd.	Information	BBB
Companhia de Bebidas das Americas	Consumer	BBB	Wipro Technologies Ltd.	Information	BBB
Companhia Vale do Rio Doce	Natural Resources	BBB+			
PetroBras: Petroleo Brasileiro S.A.	Natural Resources	BBB			
Votorantim Participações S.A.	Industrial	BBB			

Source: Standard & Poors (2008, 2009, 2010)

An initial critical hesitation might be that the firms involved in the case studies are from different sectors, thus making comparability difficult. Traditional political economy studies that focus on firms have taken a sector based approach. So, for example, a number of firms from a specific sector, such as technology, are compared across countries[44], a method which has provided political scientists with useful insights into the role of the state in sector development. Since the focal point of my study, however, is how corporations have developed and pierced in the context of a state's domestic political economy, I argue it is important to look across sectors at the broader economy. A creditworthiness approach permits me to be able to look at market drivers beyond those associated with the sector. I can then test whether or not the most successful firms reflect the comparative institutional advantages that exist in the economy, as defined by VoC theory. An ability to work across industries and in the macroeconomy, therefore, is a methodological advantage of utilizing creditworthiness as an empirical tool.

To view the piercing firms in Brazil and India from a financial or economic perspective leads to an expectation that these firms have either achieved stable and diversified cash flows or strong ties to the state. The piercing of the two Brazilian natural resource firms, CVRD and PetroBras, is easily reconciled with financial expectations of piercing. Similarly, the international ownership of AmBev and the easy access to international currencies in the businesses[45] of Aracruz, Votorantim, and Reliance are initially explained by financial theories of piercing. The remaining firms, however, are unaccounted for. Furthermore, deeper analysis of CVRD, PetroBras, AmBev, Aracruz, Votorantim, and Reliance demonstrate irregularities that cannot be ignored.

Financial theories of piercing fail to conceptualize a robust political explanation of the state-firm relationship. The fact that Bradesco and Itaú pierce run counter to the finding by Peter & Grandes that financial institutions

ISBN 978-0-9566139-1-2

should be tied tightly to the state and should not pierce. Even more confounding is the fact that the largest bank in the country by assets,[46] which is a state sponsored bank (Banco do Brasil), does not pierce,[47] since that would be the bank expected to pierce because of its direct ties to the state. In India, apart from Reliance Industries Limited, the piercing firms are all in the technology sector. Existing theory would predict that technology service firms are somehow directly tied to the state, but this is certainly not the case as all the piercing firms are privately owned with no government ownership. Dissecting the linkages and differentiations between the firm and the state is essential to understanding the forces that drive piercing.

The Indian technological outsourcing sector provides an interesting exposé about the unknown unknowns, as currently conceptualized, in the state-firm relationship and whether or not these firms are direct beneficiaries of state intervention. Certainly outsourcing is, in some ways, strategic to the Indian economy. From a global marketing standpoint, such firms represent the new India that the government is hoping to sell. Indian outsourcing in 2007-8 comprised at least 10% of exports.[48] In a broader context, however, the strategic value of outsourcing is diminished. In a country where the export to GDP ratio is roughly 23%[49] and where outsourcing comprises less than 1% of GDP[50] and employment,[51] the centrality and strategic importance of outsourcing firms in the Indian economy is questionable. Furthermore, outsourcing firms are not central to India's ability to access hard currencies, since the government can easily raise USD and EUR on international commodity markets through its many state owned natural resource enterprises. Furthermore, from a financial standpoint, the Indian government does not really need much foreign currency since it relies heavily on domestic Rupee capital markets.[52] Even the Indian government recognizes that IT is not the most important service industry in the country.[53] From a macroeconomic perspective, it is clear that India's political economy continues to primarily be focused on trying to feed its people[54] – and not providing technology solutions to the developed world. The aforementioned conditions, the limited strategic value of Indian outsourcing in the domestic economy and indirect state ties suggests that piercing should not occur.[55] Further analysis is required to understand the ties between the state and the technology sector and how such levels of creditworthiness have been achieved.

The remaining Indian firm, Reliance Industries, also does not fit well into existing financial piercing theory for a variety of reasons. While the strategic importance of Reliance to the Indian economy and its status as a natural resource firm is easily recognized,[56] why it pierces, and numerous other natural resource firms with close ties to the state do not pierce, is puzzling. The Fortune 500 list of the world's largest corporations includes five petroleum and petrochemical firms from India, and four of the firms are government owned (Indian Oil, Bharat Petroleum, Hindustan Petroleum, and Oil & Natural Gas India).[57] The four government-owned firms have obvious ties to the state and should theoretically pierce, but it is the fifth natural resource firm, Reliance, which is run as an independent, private corporation and is not tied to the state, that pierces. The contradictory piercing reality in India, even more explicitly than in the Brazilian banking scenario, illustrates the weak conceptualization of the political dimension of existing financial piercing theory.

From the perspective of the remaining Brazilian piercing firms that superficially fit into financial and economic piercing expectations, major holes remain in understanding the relationship between the firm and the state. Financial piercing theory would suggest strong ties between CVRD and the government of Brazil. The reality, however, is that CVRD was privatized in 1997, trades freely in the international market, and adheres to market standards of transparency. The government does retain a position in CVRD, but that role is limited to a uniquely structured veto stake that does not affect operations.[58] PetroBras is in a similar situation. Although it was never fully privatized, it operates much like a private corporation and even has a minority ownership stake that is freely traded in international markets.[59] With regards to AmBev, what is significant is not what is explained by financial piercing theory but what is not explained – the process of developing creditworthiness. AmBev is interesting because it is a firm that voluntarily merged with Interbrew to form InBev in 2004 in an effort to gain international market share.[60] Unlike many other emerging market corporations, the financial security provided through international ownership was driven by a merger of equals and not a sale, which accentuates the fact that this firm had achieved important status in the Brazilian consumer market before voluntarily restructuring itself. CVRD, PetroBras, and AmBev demonstrate how even when ties to the state or international guarantors/collateral exist, an unexplained dimension remains and political economic analysis is required.

A final puzzling aspect of financial piercing theory becomes apparent when viewing Brazil and India from a comparative perspective. Why is not true that the same kinds of firms pierce in each economy? Peter & Grandes suggest, for example, that financial firms are unlikely to pierce and industrial firms are likely to pierce, making

ISBN 978-0-9566139-1-2

explicit an argument presented by Durbin & Ng where firms with products that can be sold in international commodity markets should generally be able to pierce. The kinds of firms that are piercing in Brazil and India do not fit expectations. In the Brazilian case, piercing firms include financial institutions (and consumer product firms if AmBev is included). In India, on the other hand, piercing firms are outsourcing technology and service businesses; financial institutions as well as numerous natural resources firms are left out of the piercing club. The inconsistent sectors within which piercing occurs is the most explicit demonstration of how political economic theory is required.

The shortcomings of financial piercing theory help contextualize the broader ramifications and theoretical shortcomings that can be improved upon through political economy. The relationship between the firm and the state and how firms achieve success in domestic economies are just two considerations that would benefit from a new approach. Existing studies have focused too narrowly on financial explanations of piercing. Understanding of piercing was advanced through financial and economic studies, but critical analysis suggests further research, specifically in a political economy context, is required to conceptualize a coherent theory of piercing that accounts for existing empirical anomalies. And, indeed, VoC methodology has begun to successfully identify causal mechanisms in firm development.

CONCLUSION

The paper presents a puzzling reality in financial markets. The empirical link between creditworthiness and the state-firm relationship is clearly established and is then used to demonstrate why a political economy approach is required to understand firm creditworthiness and piercing. I propose that identification of the most successful firms can be better explained through VoC theory because of its focus on the state-firm relationship. While I have not yet detailed explicit empirical examples of political economic explanations, I did demonstrate the theoretical relevancy. VoC theory provides not only a methodological foundation but a coherent explanation of how a state's institutional constructions provide competitive advantages to certain firms. The empirical dimension of the paper demonstrated how existing financial and economic conceptualizations of piercing insufficiently explain the kinds of firms that achieve such levels of creditworthiness. I contend that further political economic exploration of the piercing market phenomenon will yield important results that will explain not only which firms are most likely to achieve the highest levels of creditworthiness in an economy but also provide insights into the processes of economic development.

REFERENCES

[1] The author would like to thank Dr. Lauren Phillips for extensive comments on previous versions of this paper and Susan Mase for adding style and form to previous drafts.

[2] Peter, M. and Grandes, M. 2005. How Important is Sovereign Risk in Determining Corporate Default Premia? The Case of South Africa. *IMF Working Paper*. WP/05/217.

Borensztein, E., Cowan, K., and Valenzuela, P. 2007. Sovereign Ceilings 'Lite'? The Impact of Sovereign Ratings on Corporate Ratings in Emerging Market Economies. *IMF Working Paper*. WP/07/75.

[3] See for example:

Block, S. and Vaaler, P. 2004. The Price of Democracy: Sovereign Risk Ratings, Bond Spreads, and Political Business Cycles in Developing Countries. *Journal of International Money and Finance* 23 (2004): 917-46.

Santiso, J. and Martinez, J. 2003. Financial Markets and Politics: The Confidence Game in Latin American Emerging Economies. *International Political Science Review* 24 (3): 363-395.

Bernhard, W. and Leblang, D. 2002. Political Parties and Monetary Commitment. *International Organization*. 56 (4): 803-830.

Balkan, E. 1992. Political Instability, Country Risk, and Probability of Default. *Applied Economics* 1992 (24): 999-1008.

[4] Brewer, T. L. and P. Rivoli. 1990. Politics and perceived country creditworthiness in international banking. *Journal of Money, Credit and Banking* 22(3): 357-369.

[5] Standard & Poors (S&P). 2006. *Sovereign Credit Ratings: A Primer*. New York: Standard & Poors.

[6] Standard & Poors (S&P). 2008. *Corporate Ratings Criteria*. New York: Standard & Poors.

[7] Standard & Poors (S&P). 2006.

Moody's. 2008. *A Guide to Moody's Sovereign Ratings.* New York: Moody's.

Citron, J.-T. and G. Nickelsburg. 1987. Country Risk and Political Instability. *Journal of Development Economics* 25: 385-392.

Brewer, T. L. and P. Rivoli. 1990.

[8] Standard & Poors (S&P). 2008. *Corporate Ratings Criteria.* New York: Standard & Poors.

Moody's. 2007. Global Business & Consumer Service Industries. *Moody's Corporate Finance.* New York: Moody's.

Vasicek, O.A. 1984. Credit Valuation. KMV Corporation (now Moody's KMV).

Edward, A. and Saunders, A. 1996. Credit Risk Measurement: Developments over the Last 20 Years. *Working Paper Series Fin 96-3.* New York: New York University, Stern School of Business Finance Department.

[9] Pratt, S. 1998. *Cost of Capital: Estimation and Applications.* New York: John Wiley & Sons.

[10] Langohr, H. and Langohr, P. 2008. *The Rating Agencies and their Credit Ratings: What They Are, How They Work, and Why They Are Relevant.* West Sussex: John Wiley & Sons.

[11] Chisolm, A. *An Introduction to International Capital Markets: Products, Strategies, Participants 2nd ed.* London: Wiley.

[12] One debate that is not addressed in this paper relates to whether or not credit ratings lead or lag credit spreads in the market. Ratings have a high demand in the marketplace because of information asymmetries between issuers and lenders; this suggests that ratings lead spreads. Research has confirmed this view by highlighting the systemic importance of credit rating agencies in the marketplace.

Langohr, H. and Langohr, P. 2008. *The Rating Agencies and their Credit Ratings: What They Are, How They Work, and Why They Are Relevant.* West Sussex: John Wiley & Sons.

Bruner, C. and Abdelal, R. 2005. To Judge Leviathan: Sovereign Credit Ratings, National Law, and the World Economy. *Journal of Public Policy* 25 (2): 191-217.

Sinclair, T. 2005. *The New Masters of Capital: American Bond Rating Agencies and the Politics of Creditworthiness.* Ithaca: Cornell University Press.

[13] Ederington, L., Roberts, B., and Yawitz, J. 1987. The Informational Content of Bond Ratings. *Journal of Financial Research.* 10 (3): 211-226.

[14] The local versus foreign currency distinction is an important one that will be revisited later in the chapter. To briefly summarize, theories initially predicted it would be impossible for a country to default in local currency debt, since the government al ways has the power to print more money. In reality, however, not all states have an AAA or Aaa local currency rating because rating agencies account for greater systemic costs within the domestic political economy that may be associated with a local currency default.

[15] Furthermore, if a firm were to somehow achieve credit superiority in a globalized economy, the expectation would be for that firm to abandon its home state of incorporation for a more stable operating environment.

[16] Fuchs, D. D. 2005. *Understanding Business Power in Global Governance.* Baden-Baden: Nomos.

[17] Durbin, E. and Ng, D. 2005. The Sovereign Ceiling and Emerging Market Corporate Bond Spreads. *Journal of International Money and Finance.* 24 (4): 631-649.

Moody's 2005. Piercing the Country Ceiling: An Update. *Moody's Rating Methodology.* New York: Moody's.

[18] The Dominican Republic, Ecuador, Pakistan, Russia, Ukraine, and Uruguay are provided as specific examples, by Moody's, where the government was in default yet domestically headquartered firms did not suffer from a debt moratorium that would inhibit foreign currency debt payments. See Moody's 2005.

[19] De La Plaza, L. and Sirtaine, S. 2005. An Analysis of the Uruguay Banking Crisis. *World Bank Policy Research Working Paper 3780.*

[20] Durbin, E. and Ng, D. 2005

[21] Ibid.

[22] Peter, M. and Grandes, M. 2005. How Important is Sovereign Risk in Determining Corporate Default Premia? The Case of South Africa. *IMF Working Paper.* WP/05/217.

[23] Ibid. p. 37.

[24] Ibid.

Peter, M. 2002. Estimating Default Probabilities of Emerging Market Sovereigns: A New Look at Not-So-New Literature. *HEI Working Paper* No. 6. Geneva: Graduate Institute of International Studies.

[25] Ferri, G. Liu, L., and Majnoni, G. 2001. The Role of Rating Agency Assessments in Less Developed Countries: Impact of the Proposed Basel Guidelines. *The Journal of Banking & Finance* 25 (1): 115-148.

[26] Ferri, G. and Liu, L. 2003. How Do Global Credit-Rating Agencies Rate Firms from Developing Countries? *Asian Economic Papers* 2 (3): 30-56.

[27] Borensztein, E., Cowan, K., and Valenzuela, P. 2007. Sovereign Ceilings 'Lite'? The Impact of Sovereign Ratings on Corporate Ratings in Emerging Market Economies. *IMF Working Paper.* WP/07/75.

[28] Identification and measurement of sovereign risk is extremely challenging in an advanced country context because of how little risk is generally associated with the debt of governments that have the most developed economies. For instance, US Treasuries are considered to be the closest proxy of a risk free rate of return, yet piercing has occurred in the US; econometric analysis would be unable to correctly analyze sources of risk in these situations. The piercing

ISBN 978-0-9566139-1-2

phenomenon is not only most measurable in an emerging market context, but it is also most frequent and stable in these markets.

[29] In order to test potential firms for the case study in this research, I conducted preliminary tests, approximating the models used by Peter & Grandes and Durbin & Ng to identify the explanatory value of sovereign risk relative to other financial metrics in piercing scenarios. Controlling for financial factors, I find that sovereign risk has the greatest effect on the spread of corporate debt. My model needs further development and, as it currently stands, is essentially a replica of the Peter & Grandes model with some minor changes.

[30] Durbin, E. and Ng, D. 2005.

[31] Porter, M. 1990. *The Competitive Advantage of Nations.* London: Collier MacMillan.

Dunning, J. 1997. Governments and the Macro-Organization of Economic Activity: A Historical and Spatial Perspective. In *Governments, Globalization, and International Business.* Edited by John Dunning. Oxford: Oxford University Press.

[32] Williamson, J. 1990. What Washington Means by Policy Reform. In *Latin American Adjustment: How Much Has Happened?* Edited by John Williamson. Washington D.C.: Institute for International Economics.

[33] See Stiglitz, J. 2002. *Globalization and Its Discontents.* New York: W.W. Norton & Co.

Sachs, J. 2006. *The End of Poverty: Economic Possibilities for Our Time.* New York: Penguin.

Rodrik, D. 2007. *One Economics, Many Recipes.* Oxford: Princeton University Press.

[34] Prasad, E., Rogoff, K., Wei, S., and Kose, M. 2003. Effects of Financial Globalization on Developing Countries: Some Empirical Evidence. *International Monetary Fund Occasional Paper 220.*

Prasad, E. Rajan, R., and Subramanian, A. 2007. Foreign Capital and Economic Growth. *NBER Working Paper 13619.*

[35] Buthe, T. and Milner, H. 2008. The Politics of Foreign Direct Investment into Developing Countries: Increasing FDI Through International Trade Agreements? *American Journal of Political Science* 52 (4): 741-762.

[36] Simmons, B. and Elkins, Z. 2004. The Globalization of Liberalization: Policy Diffusion in the International Political Economy. *American Political Science Review* 98 (1): 171-89.

[37] Hall, P. and Soskice, D. 2001. An Introduction to Varieties of Capitalism. In *Varieties of Capitalism: The Institutional Foundations of Comparative Advantage.* Edited by Peter Hall & David Soskice. Oxford: Oxford University Press.

[38] Milgrom, P. and Roberts, J. 1992. *Economics, Organizations, and Management.* London: Prentice Hall.

[39] Langlois, R. and Robertson, P. 1995. *Firms, Markets, and Economic Change: A Dynamic Theory of Business Institutions.* London: Routledge.

[40] Hall, P. & Soskice, D. 2001.

[41] See Hall and Soskice 2001.

[42] Feldmann, M. 2007. The Origins of Varieties of Capitalism: Lessons from Post-Socialist Transition in Estonia and Slovenia. In *Beyond Varieties of Capitalism: Conflict, Contradictions, and Complementarities in The European Economy.* Edited by Bob Hancké, Martin Rhodes, and Mark Thatcher. Oxford: Oxford University Press.

For other examples, see the range of chapters in Hall and Soskice 2001 or Hancké et al. 2007.

[43] Standard & Poors (S&P). 2008, 2009, 2010. *Corporate and Counterparty Credit Ratings that Exceed the Sovereign's Rating.* New York: Standard & Poors, May 2008.

[44] Evans, P. 1995. *Embedded Autonomy: States and Industrial Transformation.* Princeton: Princeton University Press.

Sally, R. 1995. *States and Firms: Multinational Enterprises in Industrial Competition.* London: Routledge.

[45] The listed firms are all commodity-linked and are therefore easily able to secure USD. Commodities are easily sold in highly liquid international commodity markets.

[46] Records from the Banco Central do Brasil indicate that Banco do Brasil has total assets of $397,408,923 compared to $336,321,559 of the second largest bank, Itau. The third bank, by assets, also pierces, and it only has $255,224,430 in balance sheet assets.

[47] Banco Central do Brasil. 2009. *Top 50 banks by total assets less brokerage.* www.bcb.gov.br.

[48] Indian software and services exported $40.3bn in 2007/8. Total merchandise and commercial services exports were roughly $197bn in FY2008.

Business Portal of India. 2010a. "Information Technology and IT Enabled Services." Accessed March 31, http://business.gov.in/Industry_services/information_tech.php.

World Trade Organization. 2010. *Trade Statistics: India.* Accessed March 31, 2010, http://stat.wto.org.

[49] The ratio is taken from World Bank Data 2009.

[50] The Ratio is outsourcing to total GDP. See Business Portal of India 2010 and WTO 2010.

[51] 1.63MM employed in outsourcing in 2006/7 out of a workforce of approximately 496MM in 2006. See Business Portal of India 2010 and

Department of State. 2006. *Foreign Labor Trends Report: India 2006.* Accessed March 31, 2010, http://www.state.gov/g/drl/rls/77593.htm.

[52] Outstanding Indian Securities are denominated principally in Rupees. See Bloomberg data resources and

Reserve Bank of India. 2010. *List of Government of India Securities Outstanding as on February 8, 2010.* Accessed March 31, 2010, http://www.rbi.org.in/Scripts/bs_viewcontent.aspx?Id=1956.

[53] Business Portal of India. 2010b. "Services." Accessed March 31, http://business.gov.in/Industry_services/services_sector.php.

ISBN 978-0-9566139-1-2

[54] Anand, G. 2002. Green Revolution in India Withers as Subsidies Backfire. *Wall Street Journal*, February 24, Europe Edition.

Kazmin, A. 2010. Food Crisis Worsens Amid Call for 'Green' Reform. *Financial Times*, February 10.

Lamont, J. 2010. New Delhi to Reign in Subsidies. *Financial Times*, February 24, World News.

[55] One financial response would be that these firms pierce because of their international cash flows, but then the question as to why these firms achieved such creditworthiness remains. Why should technological outsourcing firms pierce and not other exporters. Furthermore, what explains the rise of the IT industry and their ability to achieve capital account carve outs? A political dimension must be explored to answer these questions.

[56] Reliance Industries comprises 10.4% of India's exports and 6.1% of Total Market Capitalization in India. See Reliance Industries Limited. 2009. *Annual Report 2008-2009*.

[57] Fortune. 2009. *Global 500 Corporations*. Accessed April 15, 2010, http://money.cnn.com/magazines/fortune/global500/2009/.

[58] As of FY2008, the government retains certain veto rights in the operation of CVRD. The government also has an equity stake in the company, directly owning 2.9% of preferred shares as well as 6.7% of preferred shares through the Banco Nacional do Desenvolvimento. The government, through pension funds, indirectly owns a majority stake of common shares, but this ownership is viewed as passive since the shareholders are individual citizens.

Companhia Vale do Rio Doce. 2009. *Annual Report on Form 20-F 2008*. Retrieved from: http://www.vale.com/vale_us/media/20F_2008_i.pdf

[59] As of FY 2008, the Brazilian government owns 55.7% of outstanding common shares of PetroBras; common shares are the only voting shares (Petróleo Brasileiro S.A. 2009). Petróleo Brasileiro S.A. 2009. *Annual Report on Form 20-F 2008*. Retrieved from: http://www2.petrobras.com.br/ri/spic/bco_arq/Form20-F2008FinalIng.pdf

[60] AmBev. 2004. "Interbrew and AmBev Complete Combination to Establish InBev". AmBev Press Release, August 27, 2004.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

'NO CASTLE-LIKE MAGNIFICENCE' – WILLIAM FAULKNER AND THE SOUTHERN PLANTATION MANSION

Dorette SOBOLEWSKI

MLitt Student, University of Glasgow, Glasgow, Scotland
d.sobolewski@freenet.de

Keywords: US Southern Literature, Plantation Novel, Architectural History, Southern Aristocracy

ABSTRACT

William Faulkner follows one tradition of his Southern literary predecessors by locating the plot of his novel *Absalom, Absalom!* in a plantation mansion, a common setting in Southern literature. Analyzing the mansion in its role as a literary setting may help to provide an understanding of how Faulkner selectively incorporates the historical realities of Southern mansions into his work. Over and beyond his use of the mansion as a common setting, Faulkner's utilisation of that mansion functions as more than a mere setting – it acts as a symbol as well. For that matter, if one is to understand Faulkner's work in a Southern context, it is crucial to understand both how Southern plantation mansions in literature can serve as symbols, and what they in effect symbolize. This requires exploring conventional depictions of a Southern plantation mansion in Southern fiction, comparing them with the findings of historians, and then relating these to Faulkner's fictitious application.

An interdisciplinary approach to Faulkner's fiction based on contextualisation to other Southern fiction and architectural history can provide rewarding research that overcomes the restrictions of traditional literary theory. I will show how selected major Southern writers prior to Faulkner depicted the Southern mansion in their works, such depictions subsequently becoming an integral element in the literary tradition which Faulkner later subscribed to. Of importance here is the question to what degree those writers' fictitious portrayals reproduces historical realities of the plantation mansion. Following the line of this topic, I will also examine the extent to which William Faulkner's depiction of the plantation mansion corresponds to the traditional role which such an edifice plays in Southern literature. Such an interdisciplinary approach functions effectively to utilize the Southern mansion in Faulkner's work to resolve the difference between literary and historical perspectives on that Southern icon.

INTRODUCTION

William Faulkner's fiction is born out of a distinct tradition developed in the writings of earlier Southern authors who reflected on and/or interpreted Southern society in their works. As a common setting and the focal point of their works, such authors often chose the locus of the Southern plantation.[1] A representative example of that literary tradition can be found in Ellen Glasgow's description of Blake Hall in her 1904 novel, *The Deliverance*:

> *[T]he house presented a cheerful spaciousness of front – a surety of light and air – produced in part by the clean white Doric columns of the portico [...] From the smoking chimneys presiding over the ancient roof to the hospitable steps leading from the box-bordered walk below.*[2]

In *The Deliverance*, Ellen Glasgow names her fabricated plantation Blake Hall after the fictitious original owner of that plantation. In *Absalom, Absalom!* Faulkner similarly gives his imaginary plantation, Sutpen's Hundred, the

ISBN 978-0-9566139-1-2

name of its fictional first owner, Thomas Sutpen. In these two instances, then, both authors paid similar allegiance to the Southern historical custom of often naming plantations after their founding owners.

In his actual description of Sutpen's Hundred, however, Faulkner differs strikingly from Ellen Glasgow's. In contrast to her detailed depiction of Blake Hall, Faulkner never really describes his mansion, but instead only provides snippets of information about it. He specifically refers to it as a 'mansion,'[3] but says nothing more about it than that it was 'a house the size of a courthouse'.[4] This lack of detail about this fictitious mansion building is not unusual in Faulkner's writing; he uses a similar style in the short story 'Barn Burning'. There, the mansion of Major de Spain is seen through the eyes of the ten-year-old Colonel Sartoris Snopes, the son of Poor Whites who move from plantation to plantation in search of work. Faulkner's only comment at the time Snopes first encounters de Spain's mansion is that 'he had never seen a house like this before. "Hit's big as a courthouse".'[5] This brief observation is as lacking in detail as Faulkner's description of the mansion at Sutpen's Hundred in *Absalom*. In both cases, Faulkner depicts the scale of the mansion by comparing it to a courthouse, but provides no further information about the architectural style or any other features.

These two examples from Faulkner's writings demonstrate a deliberate vagueness in his descriptions of the narrative's time-honoured central focal point, the mansion house. On the one hand, he conforms to the earlier tradition in Southern literature of choosing a plantation mansion as the locus of his work. At one and the same time, though, he breaks with that tradition by rejecting the explicit descriptions of the plantation mansions characteristic of his Southern writer predecessors in their works. When combined, these factors demonstrate a stylistic selectivity or inconsistency which create a feeling of ambivalence in Faulkner's use of the plantation mansion as a literary location.

It cannot be taken for granted, however, that Faulkner's divergence from this literary tradition is simply a question of style. Instead, it seems more likely that it results from a different, more complex aspect of Faulkner's entire approach to his literature, namely his strong emphasis upon the truthful rendition of "historical realities". Faulkner's respect and striving for historical reality is perhaps best exemplified by the way in which he addressed the racial and class differences that were part of life in the South at the time of his writing. This fictional yet factual presentation of historical realities can perhaps best be characterized as a distinctive Faulknerian form of "literary history".

Literary history can be generally defined as an approach to the creation of history-related literature which accurately reflects the history which it is referencing. Obviously, there is an inherent duality to the concept itself, and Lee Patterson indicates one source of that duality in his definition of literary history:

> *On the one hand, its commonsense meaning refers it to an immanent or intrinsic history of literature [...] Yet the term also describes a critical practice concerned not with the history of literature as a self-contained cultural activity but with the relation of literature, as a collection of writings, to history, as a series of events.*[6]

Patterson further argues that literary history itself can be seen as either a depiction of or an influence on history – or, indeed, even as a perception of history. He proposes that a text extracted from a specific historical context can, on the one hand, be based on historical realities and, at the same time, become the source for an interpretation of that history comprising those realities. As the renowned Faulkner scholar Cleanth Brooks states, a bi-directional approach such as Faulkner's (e.g. both drawing on historical realities and simultaneously referencing their frequent influences on and reflections through subsequent literary tradition) can result in the advantage that '*such reading may well confirm the reader's confidence in the penetration of Faulkner's insight into the culture that produced him*'.[7] On the other hand, though, literary fiction – including Faulkner's – can easily be promoted as possessing an authority on the history it deals with, regardless of the degree to which it remains loyal to whatever the historical realities might have been (or assumed). Such assumed (or self-assumed) historical authority can, of course, be easily abused, and consequently an element of caution is recommended against taking any literary history at face value – including Faulkner's. Obviously, such potential pitfalls of literary history must be given due consideration when analyzing Faulkner's depictions of the Southern plantation mansion. An interdisciplinary approach can provide the means for such due consideration; in this case a comparison of Faulkner's fictitious descriptions of the plantation mansion and the historical realities of those mansions as present in historical sources.

ISBN 978-0-9566139-1-2

When addressing Faulkner's novel *Absalom, Absalom!*, Cleanth Brooks provides a realistic analysis of what such an interdisciplinary examination should lead to:

> [*W*]*hat some of the more recent historians of the South depose may throw needed light on the novel, and may give us more confidence that Faulkner, in writing about his Yoknapatawpha planters, was not simply indulging in myths and legends of the Old South that had no real reference to the history of the society in which he grew up.*[8]

It must be noted that Brooks' research took place in the years immediately after the Civil Rights movement of the 1960s. Fortunately, many journals and magazines have been digitalized in the past few years, and more cultural studies have been reprinted. As a result, it is now possible to extend historical research back to the 1920s and '30s and encompass sources which were not readily available at the time Brooks was writing. The incorporation of articles by historians contemporary to Faulkner can certainly help to shed light on the '*relation of literature, as a collection of writings, to history, as a series of events*,' as Patterson calls it. Accordingly, this paper now turns to an examination of the degree to which Faulkner, while complying to the Southern literary traditions regarding the plantation mansion as a narrative setting, subjected those same traditions to the scrutiny of his own perception of historical reality and modified them accordingly. I will also examine the extent to which William Faulkner's depiction of the plantation mansion corresponds to the traditional role which such a structure played in Southern literature. Finally, through an interdisciplinary approach I will attempt to evaluate the degree of historical accuracy about those structures in both Faulkner's and his predecessors' works.

DISCUSSION

Initially, it is necessary to analyze how selected Southern writers prior to Faulkner depicted the Southern mansion in their works, works which subsequently become integral elements in the literary tradition which Faulkner later subscribed to. Of central importance here is the question of the degree to which those writers' fictitious portrayals reflected the historical realities of the plantation mansion. Before entering into a description of how selected Southern writers prior to Faulkner depicted the Southern mansion though, it is important to address a complex of traditions and historical developments which led to a phenomenon which I will refer to as 'Southern literary ambivalence' and which came to the fore during the Southern Renaissance of the 1920s and '30s. Even though this concept might seem concomitant to some of the dualities and ambiguities dealt with in the introduction, it is, by itself as well, a crucial element for the critical analysis of William Faulkner's work.

The iconic historian of the South, C. Vann Woodward, critically noted how Southern tradition and historical awareness persisted well into the time of the Southern Renaissance. '*The South had its full share of illusions, fantasies, and pretensions, and it has continued to cling to some of them with an astonishing tenacity that defies explanation*'.[9] Richard H. King adds to this by making clear not only how tenacious such elements of Southern culture were, but also how that tenacity created a significant dichotomy of attitudes and feelings among Southern Renaissance authors. This dichotomy manifested itself in the common 'Southern literary ambivalence':

> *[T]he writers and intellectuals of the south after the late 1920s were engaged in an attempt to come to terms not only with the inherited values of the Southern tradition but also with a certain way of perceiving and dealing with the past, what Nietzsche called "monumental" historical consciousness. It was vitally important for them to decide whether the past was of any use at all in the present; and, if so, in what ways? Put another way, the relationship between present and past which the Renaissance writers explored was fraught with ambivalence and ambiguity[...].*[10]

Tradition and history were not the only factors impressing heavily upon Southern writers at the time. The 1930s saw an attempt by some white Southerners to recreate an image of society as it had been during the years of slavery – something that many white Southerners from the generations since the Civil War had considered desirable. Also in the 1930s, the Great Depression gave to the South the questionable privilege of becoming '*the Nation's No. 1 economic problem*'.[11] The despair and poverty which the South experienced at that time made the old order of the antebellum Southern society, Old Dixie, seem even more desirable (at least from a White perspective). This feeling of nostalgia often manifested itself in attempts to recreate that society and was frequent topic in the literature of the Southern Renaissance. That was a society that was ordered and organized – but also one that enslaved people. It was a society that was fondly remembered – and at the same time hated. And, it was a society glorified in memory – but one destroyed by defeat. It does not require too much imagination to picture the culturally and historically astute Faulkner as sensitive to the pangs of such intense conflicts, both without and within himself. A reference to such turmoil can be found in the prologue to the

ISBN 978-0-9566139-1-2

biography on William Faulkner in the Norton Anthology of *The Literature of the American South*. Here it is claimed that Faulkner experienced a hotchpotch of feelings in regard to his heritage: '*feeling pride in his ancestors and his region but also burdened with the sins of the fathers [e.g. slavery]*'.[12]

The resulting feelings of ambivalence toward his home region would certainly not have been unique to Faulkner, and, like many of his white contemporaries from the South, he articulates such feelings in his work. His convincing portrayal of the character Quentin Compson in *Absalom* indicates more than a passing acquaintance with such emotions. Plagued by frequent attacks of such emotional conflict, Compson repeatedly has to reassure himself that he does not hate the South: '*I don't. I don't! I don't hate it! I don't hate it!*'[13]

The roles which the various plantation mansions play in the lives of Faulkner's characters such as Thomas Sutpen in *Absalom*, or Abner Snopes, the Poor White antagonist in Faulkner's short story '*Barn Burning*', show the importance of those mansions as literary vehicles. Faulkner's 'Southern literary ambivalence' would understandably influence his use of such a major component of his work. One does not have to search too far to find manifestations of such influence.

James C. Bonner, a contemporary of William Faulkner's, commented on the importance of how a building can reflect on life in and around it:

> *The houses in which people have lived in other periods of their history were intimately a part of the whole structure of their lives; they reflect the social and cultural aspects of the period. An understanding of those buildings will contribute to an understanding of the life which took place around them.*[14]

Bonner clearly suggests that through the analysis of people's homes, conclusions can be drawn on the society which such home owners occupied. However, he mainly refers to the actual buildings themselves. Of at least equal importance is an understanding of how expected behaviour patterns in and around the houses can likewise be seen as manifestations of 'social and cultural aspects of the period'. In regard to the Southern plantation mansion, the interaction of literature and history provides a pivotal basis for such understanding.

Shortly after John Pendelton Kennedy introduced the plantation novel with *Swallow Barn* in 1832, it became and has remained ever since a subgenre in Southern fiction. Further, the plantation – and at its center the master's mansion – has continued to be one of the most dominant settings in the literature of the American South. The Meriwether clan in Kennedy's novel does not live in a grand classical building with Doric columns, but rather in a somewhat piecemeal house that, even though it has seen better times, still demonstrates a unique character which obviously makes it dear to its inhabitants.

> *The buildings illustrate three epochs in the history of the family. The main structure is upwards of a century old; one story high, with thick brick walls, and a double-faced roof, resembling a ship, bottom upwards; [...] To this is added a more modern tenement of wood, which might have had its date about the time of the revolution: it has shrunk a little in the joints, and left some crannies, through which he winds whisper all night long. The last member of the domicile is an upstart fabric of later times, that seems to be ill at ease in this antiquated society[...].*[15]

This kind of description is by no means unique to Southern literature. What is more, it indicates that, with certain qualifications as to time and place, Southern planters may actually have lived in rather mediocre, non-descript, eclectic structures which were altered and enlarged whenever the need might arise.[16] Indeed, research into historical sources on this topic gives a degree of historical validity to Kennedy's depiction. Charles Bonner argues that, only in the 1850s, did the rural South begin '*to formulate a philosophy of rural architecture*',[17] from which a tendency towards the classical Greek design originated. Furthermore, Francis Pendleton Gaines argues in his 1924 study *The Southern Plantation* that there neither was nor is an abundance of the glorious mansions so commonly associated with the South: '*The South was no more thick-set with elegant mansions than Elizabethan England was crowded with Kenilworth castles*'.[18] Through the familiarity of their grandeur, George Washington's Mount Vernon, Thomas Jefferson's Monticello,[19] or Robert E. Lee's Arlington may seem to imply a wide-spread incidence of such glorious antebellum plantation mansions. However, Bonner, in concurrence with Gaines, marks this as a misconception: '*A great house of classic design in a rural setting was a rare phenomenon, even among the more prosperous planters*'.[20] A major reason behind this misconception of the ubiquity of such mansions can be found in the influence of the arts on concepts of historical realities.

ISBN 978-0-9566139-1-2

> *A thousand stories, stage "sets," or motion pictures, all assuming a Mount Vernon or a Montpelier, lead eventually to the conclusion that there must be tens of thousands of such mansions, placed conveniently a few miles apart on every public road.*[21]

This is a significant example of what Patterson refers to in his essay on '*Literary History*': the interaction of literature and history. Because many novels, plays, and movies produced after the defeat in the Civil War in 1865 have so long and frequently utilized the old Southern mansion to celebrate the Lost Cause and utilized them as such an intrinsic setting (perhaps most prominently in both the 1936 novel and the 1939 movie adaptation of Margaret Mitchell's *Gone With the Wind*), people have come to believe such presentations as the manifestations of a factual omnipresence of such mansions. This is where an interdisciplinary analysis becomes invaluable, for it reveals such discrepancies between reality and fiction. Gaines goes further and compares the depiction of the Southern mansion in literature with the actual homesteads of planters from these regions. Interestingly, he notes – like Bonner in regard to the actual development in rural Southern architecture in the 1850s – that in the literature set in the late antebellum period, a distinct and clearly defined style of architectural design was represented: '*Commonly the house is a large, two-story, rectangular edifice, of either brick or frame, with an impressive width of porch and great white columns, usually Doric, though the entablature varies'*.[22] Such a representation can be found in George W. Cable's story '*Belles Demoiselles Plantation*':

> *[...] Belles Demoiselles Mansion, with its broad veranda and red painted cypress roof [...] The house stood unusually near the river, facing eastward, and standing four-square, with an immense veranda about its sides, and a flight of steps in front spreading broadly downward, as we open arms to a child.*[23]

The story was published in 1879, and in it Cable portrays an antebellum mansion in a similar way in which the arts, at least according to Francis Pendleton Gaines, made ubiquitous the idea that the likes of Arlington and Mount Vernon were typical, almost universal representations of Southern mansions. Moreover, the *'immense veranda*' and the '*flight of steps*' described by Cable do seem to be a common motif that may have contributed to such an overgeneralization of Southern rural architecture. Over and beyond that, the simile '*as we open arms to a child*' is an effective mechanism to give the building some symbolic meaning – that of the mansion as a welcoming, embracing, almost motherly, sanctuary where guests are always welcome. Of course, this symbolism in Cable's story is negated when the mansion, and with it the future of the occupying family clan, descends into the Mississippi. Nevertheless, the symbolism inherent in the depiction of the welcoming and grand Belles Demoiselles mansion differs noticeably from the symbolic meaning which Faulkner attaches to the faceless, nondescript mansion in *Absalom, Absalom!*

Further, writers prior to the 1850s did not focus on providing highly detailed archetypical stereotypes of plantation mansions. Instead of an exact depiction, they instead tried to convey an impression, an aura of grandeur, as Gaines suggests:

> *In regard to definite details, the romances are vague. In certain of the novels of domestic sentiment, there is merely the implication of grandeur, the assurance that the house is a fitting edifice to serve as "the capitol of a small republic," There is no consistency of architectural effect but the element of significance is never lacking.*[24]

In this vein, in *Swallow Barn* John Pendleton Kennedy also implies such grandeur, even despite the fact that the house he portrays is rather run down and has seen better days:

> *It is shaded by a narrow porch, with a carved pediment, upheld by massive columns of wood sadly split by the sun. A court-yard, in front of this, of a semi-circular shape, bounded by a white paling, and having a gravel road leading from a large and variously latticed gate-way around a grass plot, is embellished by a superannuated willow that stretches forth its arms, clothed with its pendant drapery, like a reverend priest pronouncing a benediction.*[25]

The 'massive columns of wood sadly split by the sun' imply that, despite their decay, in earlier days this house provided a clear representation of the wealth and power of its owners. Just as George W. Cable ends his description of a Southern plantation mansion with a simile, so does Kennedy in his: a willow close to the mansion is described '*like a reverend priest pronouncing a benediction'*. The religious imagery seems to impart a spiritual ambience to the house and, correspondingly, divine validation of the status of its owners. It is the usage of such similes that serves to convey the implication of grandeur of Southern plantation mansions in Southern literature.

ISBN 978-0-9566139-1-2

Staying within this tradition, William Faulkner also implies such grandeur in his works, for example in his comparisons to courthouses. However, the way in which he accomplishes this is different from that of his predecessors. As already mentioned, there is a significant absence of detailed description of Sutpen's mansion. Throughout the fragmented narrative structure of *Absalom*, the house is identified as a 'mansion'[26] and as a 'big house'.[27] These characterizations certainly show the scale of Sutpen's mansion, but, similar to the earlier Southern works prior to the 1850s, they lack detail. Further, Faulkner also draws on similes to create an impression of the mansion in question. It is portrayed as possessing an implicit 'grim and castle-like magnificence'[28] and is explicitly said to be 'the largest edifice in the county, not excepting the courthouse itself'.[29] An awe-inspiring experience very similar to the one Colonel Sartoris Snopes has when he first sees the mansion is shared by young Thomas Sutpen, a boy of 'eleven or twelve or thirteen',[30] in the subplot in *Absalom:* 'the biggest house he had ever seen'.[31]

This raises the question why Faulkner departs from the post-Civil War Southern literary tradition of his predecessors by providing the reader with even more limited descriptions of Southern plantation mansions. One possible answer is that Faulkner perhaps saw any descriptive detail as redundant. It is no longer necessary to provide the reader with yet another detailed description of an archetypical Southern plantation mansion: the reader has already been exposed to such descriptions for more than a century. Further, the outside appearance is irrelevant, as such mansions are basically interchangeable. This is suggested, for example, in 'Barn Burning'. The ten-year-old Colonel Sartoris Snopes '*could see the grove of oaks and cedars and the other flowering trees and shrubs where the house would be, though not the house yet*'.[32] This implies an identical nature, almost a 'cookie-cutter' mentality, behind the structural set-up of Southern plantations which may differ in design but remain the same in sentiment and effect. Through the gypsy-like travels of his family from one plantation to another, the young boy has learned exactly where the mansion will be, even if the actual architecture of the individual structure might differ. And according to the analysis of Faulkner's *Absalom* and '*Barn Burning*', such a design is actually irrelevant. The apparent absence of descriptive passages in Faulkner fiction implies this redundancy.

CONCLUSION

The plantation mansion is the setting of the stories discussed: a symbol of wealth and an image of Southern aristocracy, the superiority of White planters. This is the sentiment that continued to throw its shadow over much of Southern society in the 1930s and '40s when Faulkner was writing *Absalom, Absalom!* and '*Barn Burning*'. Faulkner follows in the tradition of his Southern literary predecessors by locating his plot in a plantation mansion, a common setting in Southern literature. The analysis of the mansion in its role as a literary setting provides an understanding how Faulkner selectively incorporates the historical realities of the concept of Southern mansions into his work. Faulkner's utilisation of the mansion functions as more than a mere setting though – like other southern writers before him, he uses it as a symbol as well. However, instead of merely copying the symbolism of wealth and superiority prevalent in the fiction of his predecessors, Faulkner inverts their ideology of White supremacy. His non-descriptive depictions of the 'big house' in *Absalom* and '*Barn Burning*' at least imply criticism of the grandeur and superiority which the Southern Aristocracy tried to signify with their mansions. Yet in accepting the romanticism which was attached to the Old South and was manifested by the creation of the classic plantation mansion as a major aspect in the works of Southern writers, Faulkner joined his contemporaries during the Southern Renaissance in creating a probably unintentional distortion of history: the idea 'that there must be tens of thousands of such mansions, placed conveniently a few miles apart on every public road'.[33] By locating a major part of his plots at such plantation mansions, Faulkner actually contributed to such an idea. Perhaps the lack of description of the mansions in Faulkner's writing relied on this very aspect: the assumption that his readers would know what kind of mansion he was depicting, since Southern literature had provided so many stereotyped examples. In the wake of this notion it is safe to say that the mansion remains an important device in Faulkner's fiction which he uses to negate premeditated assumptions both of Southern society and its literary heritage.

An interdisciplinary approach such as this – exploring the conventional depictions of Southern plantation mansions in Southern fiction, comparing those with the architectural findings and descriptions of historians, and finally relating these to Faulkner's fictitious applications – can shed light on the discrepancies between fiction and reality. Once the existence of such discrepancies has been established, we can then address the question of the degree to which Faulkner was aware of and, indeed, intended such discrepancies. There are

ISBN 978-0-9566139-1-2

substantial indications that Faulkner's intention was to free himself from certain limitations of previous Southern literary traditions and therefore exercise criticism not just of the Southern aristocracy but also of his fellow Southern writers who helped to nurture the image of white supremacy. Perhaps his success in fulfilling this intention is one of the factors which contribute to his uniqueness among Southern authors. His feelings of ambiguity and ambivalence toward his home region obviously manifest themselves in an ambiguous and ambivalent approach in his writings on the South. The question remains: Did he hate it?

REFERENCES

[1] When one considers the fundamentally rural, decentralized, non-urban nature of the American South until well into the 20th Century, it is rather self-understood that the focal point of so much of the history of that society would be chosen as the focal point of so much of the literature about that society.
[2] Glasgow, E. *The Deliverance* (1904) New York: Doubleday, p. 15.
[3] Faulkner, W. *Absalom, Absalom!* (1936, 2005) London: Vintage Books, p. 11.
[4] Ibid. p. 16.
[5] Faulkner, W. 'Barn Burning' (1938), in William Faulkner, *Collected Stories* (1995) London: Vintage, (pp. 3-25) p. 10.
[6] Patterson, L. 'Literary History', in *Critical Terms for Literary Studies,* ed. by Frank Lentricchia and Thomas McLaughlin (1995) London: University of Chicago Press, (pp. 250-62) p. 250.
[7] Brooks, C. *William Faulkner. Toward Yoknapatawpha and Beyond* (1978) London: Yale University Press, p. 296.
[8] Ibid, p. 285.
[9] Vann Woodward, C. 'The Irony of Southern History' (1960), in William Faulkner, *The Sound and the Fury* (1929), ed. by David Minter (1994) London: Norton, (pp. 241-43) p. 242.
[10] King, R.H. 'Exploring the Southern Renaissance', in *Major Problems in the History of the American South. Volume II: The New South.* ed. by Paul D. Escott et al., 2nd edn (1999) Boston: Houghton Mifflin, (pp. 297-300) p. 297.
[11] Roosevelt, F.D. 'The President's Council Reports on Southern Economic Conditions' (1938), in *Major Problems in the History of the American South. Volume II: The New South.* ed. by Paul D. Escott et al., 2nd edn (1999) Boston: Houghton Mifflin, (pp. 320-21) p. 321.
[12] *The Literature of the American South. A Norton Anthology*, ed. by William L. Andrews et al. (1998) New York: Norton, (pp. 435-39) p. 436.
[13] Faulkner, *Absalom,* p. 378.
[14] Bonner, J.C. 'Plantation Architecture of the Lower South on the Eve of the Civil War', (1945) *Journal of Southern History*, 11 (pp. 370-388) p. 370.
[15] Kennedy, J.P. 'Swallow Barn; or, A Sojourn in the Old Dominion. Chapter I' (1832), in *The Literature of the American South. A Norton Anthology*, ed. by William L. Andrews et al. (1998) New York: Norton, (pp. 60-2) p. 60.
[16] Thomas Jefferson's mansion at Monticello is a good example for such eclecticism and bares witness to such experimentalism in Southern architecture.
[17] Bonner, p. 376.
[18] Gaines, F.P. *The Southern Plantation* (1924, 1962) Gloucester, MA: Columbia University Press, p. 168.
[19] At the time of his death Jefferson was highly indebted, because of the continuous experimental rebuilding of his mansion. After he had died, the mansion was abandoned and Monticello, today a symbol for antebellum wealth of the Southern aristocracy, fell into such state of disrepair that it almost was torn down.
[20] Bonner, p. 371.
[21] Gaines, p. 169.
[22] Ibid, p. 166.
[23] Cable, G.W. 'Belles Demoiselles Plantation' (1879), in *The Literature of the American South. A Norton Anthology*, ed. by William L. Andrews et al. (1998) New York: Norton, (pp. 277-88) p. 278.
[24] Gaines, p. 166.
[25] Kennedy, p. 61.
[26] Faulkner, *Absalom,* p. 11.
[27] Ibid, p. 15.
[28] Ibid, p. 38.
[29] Ibid, p. 39.
[30] Faulkner, *Absalom*, p. 228.
[31] Ibid, p. 227.
[32] Faulkner, 'Barn Burning', p. 10.
[33] Gaines, p. 169.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

BAYESIAN CHANGE POINT DETECTION FOR SATELLITE FAULT PREDICTION

Ryan TURNER

PhD Student, University of Cambridge, Cambridge, UK
rt324@cam.ac.uk

Keywords: Machine learning, Time series, Bayesian, Change point detection, Reliability, Statistics

ABSTRACT

Change point detection makes predictions in data whose structure changes over time and belongs to a class of methods called time series methods. Machine learning is the design of algorithms whose performance improves with data. Techniques from statistics, such as change point detection, have received increased interest in recent years in machine learning. Many problems have change point structure: Stock markets exhibit change points when some significant economic event causes an increase in volatility, weather systems may exhibit change points during years of El Nino, and electronic systems show change points when devices begin to fail or configurations change.

This paper focuses on detecting changes in the operation of satellite base stations. The aim is to provide predictions of device lifetimes or time until maintenance will be necessary. Typical reliability engineering uses survival analysis to find lifetime distributions and mean time between failures (MTBF). However, using change point detection we can make predictions that include information dynamically. For instance, a change in the measured weather conditions could signal an impending satellite signal loss due to an imminent storm. Modeling changing dependencies is also important; a change in the relationship between the external and internal temperature on a device should mean there is a problem with the cooling unit. In the mentioned examples, the weather change point occurred when the weather conditions changed and when the cooling unit began to malfunction, respectively. Unlike MTBF, change point detection provides probability distributions describing the likelihood of a failure, which is necessary in most cases since there is usually significant remaining uncertainty over the time until failure. The two key aspects are that change point detection can include information dynamically and model changing dependencies.

We show how to use techniques from Bayesian statistics, namely change point detection, to model and understand electronic systems.

INTRODUCTION

Change point models are used to extend simple time series methods to cases where the appropriate model changes. In this paper we look at modeling sensor measurements from a satellite base station using change point methods. We provide examples where events, such as power supply problems or changes in the amplifier power, correspond to change points in measurements taken by a logging system. We do this by extending the Bayesian online change point detection (BOCPD) algorithm introduced by [1].

This paper is organized as follows: In Section 2 we describe the notion of time series data and what it means for an algorithm to be online. In Section 3 we give some background on what machine learning is and how it is used. We also explain its relationship to statistics as well as how ideas from Bayesian statistics carry over to Bayesian

ISBN 978-0-9566139-1-2

machine learning. In Section 4 we explain the BOCPD algorithm both mathematically and through illustrative examples. We demonstrate the use of BOCPD on real world data, namely satellite fault logs, in Section 5.

TIME SERIES METHODS

Change point detection is designed to deal with time series data. The time series measurements we deal with are real valued and can be sampled at any point in time. Examples include voltage, temperature, and current measurements. In more precise terms time series data maps from a time index $t \in T$ to a measurement $x_t \in R$. Time series can be in discrete time ($T = Z$) or continuous time ($T = R$), which means discrete time is a special case of continuous time. For example, if we measure the voltage exactly once an hour we can model it as discrete time where the time index is how many samples have been taken so far. We call this case *uniform sampling*. However, if voltages are measured at arbitrary times or with changing sampling rates it is more convenient to model it as a continuous time process. To simplify the analysis we focus on discrete time processes.

Online Algorithms

We also assume that our application demands *online* change point detection: an autonomous robot must detect change points in its sensors as they happen, an automatic trading system must detect change points as soon as possible, and electronic monitoring applications must alert to change points in measurements as soon as possible.

The alternative to online algorithms are *retrospective* algorithms. A retrospective fault detection system would be one where it waits until the end of the year, or some other period of time, and then uses all the data throughout the year to locate where the change points are. By contrast, in the online system if a change point happens in March, the system should alert as soon as it can, and not wait till the end of the year to collect all the data it can before doing any analysis. Prior to the work by [1], and similar by [6], previous Bayesian approaches to change point detection have been retrospective [2, 14].

MACHINE LEARNING

Machine learning is the study of algorithms whose performance improves with exposure to data. The functionality of typical software can be completely specified in advance. However, in machine learning software must learn to match its output to training examples. The classic example is optical character recognizers (OCR): images of characters, say zero to nine, are provided and the software must output which digit is in the image. As opposed to trying to specify rules about what makes a two a two, example images are provided along with labels (this is known as the *training set*). A good machine learning algorithm will predict the correct character in test when novel images are provided. An OCR is an example of an *iid data set*, this paper is focused on the more complex task of *time series data*.

It was realized early on that machine learning must have closer ties to statistics than other parts of computer science; functionality was being specified by real data not abstract specifications on paper as is the case with data structures like stacks and queues. However, machine learning still maintains some independence from statistics. Machine learning is more concerned about new efficient algorithms for complex tasks while statistics is more concerned about the details of data sets. Central to machine learning is *test set* performance.

Measuring Performance

Given that machine learning is interested in predictive performance we need a way to measure the performance. There is a standard procedure used in machine learning; for simplicity, we first consider the case of iid data. For iid data, the evaluation procedure works as follows. Each data point is randomly assigned to be in either a training set of size N or a test set of size N'. It is a common novice mistake to test a model on the same data set the model was trained on. This gives overly optimistic estimates of performance and does not tell us how well an algorithm will perform on new data.

ISBN 978-0-9566139-1-2

The models' parameters θ are set in a way specified by the model designer using the training set.[1] Then on each point in the test set the model makes a prediction about the data point x_i. If the model outputs a probability distribution $p(x_i)$ we convert it to an action a (or point estimate) by minimizing the expected loss

$$a_i = \operatorname{argmin}_a \mathbf{E}[L(x_i, a)] = \operatorname{argmin}_a \int_i L(x_i, a) p(x_i)\, dx \qquad (1)$$

This optimization is Bayes' decision rule. Certain methods do not provide a predictive distribution p, and directly give an action a. If x is continuous ($x \in R$) common loss functions are mean-square-error (MSE), $L(x_i, a) = (x_i - a)^2$, and mean-absolute-error (MAE), $L(x_i, a) = |x_i - a|$. The optimal actions for MSE and MAE are to supply the mean and median of p, respectively. For the log loss or intrinsic loss, $L(x_i, p) = -\log p(x_i)$, the optimal action is to supply p itself [5]. The models' performance or empirical loss $\hat{l}$. is the average loss over each test point. See [3] for more details.

In the time series context, the model can be evaluated by its loss on the next step ahead prediction. In other words, predicting x_i given $x_{1:i-1}$. Theoretical perspectives on evaluating time series models is explored in [4], most of which was inspired by evaluating probabilities in weather prediction.

Bayesian Machine Learning

A significant cohort of machine learning researchers borrow ideas from Bayesian statistics. Bayesians are willing to use probability to express uncertainty over any well defined quantity; sometimes these probabilities are interpreted as state of belief or knowledge. This is in contrast to the frequentist alternative where only samples from "repeatable" processes are treated as random and under the jurisdiction of probability theory.

This division in statistics has consequences in change point detection. Much of the older work on change point detection, dating back to [8], was based on frequentist hypothesis testing ideas. Such setups are designed to signal an alert when a change point occurs; however, they only say the probability of an alert *given there is no change point* is less than α, usually 5% (this is known as a false positive rate). In Bayesian change point detection, we say the probability that there was a change point, *given the time series we observed*, is $x\%$. The substantive difference between the two statements is often puzzling to those not well versed in probability. The difference is made most clear by extreme case reasoning. There is no unique procedure in the frequentist change point detector, a change point detector that completely ignored the data and alerted randomly 5% of the time would satisfy the frequentist statement. There is no unique procedure in the Bayesian change point detector either; in the Bayesian change point detector the designer must first specify a prior probability distribution for what one would likely see given a change point. Once the prior assumptions are specified, there then exists a uniquely optimal procedure for finding the change points.

The division has a slightly different context in machine learning. Machine learning is concerned about test set performance, so we would like to specify algorithms that make the best predictions on new data. However, it is not possible to say before being given a data set what the best algorithm will be. The "no free lunch theorem" says that algorithms that perform better on one data set will lose some performance on others. Bayesian machine learning cares about achieving optimal performance on average under our prior assumptions. By contrast, frequentist analysis is typically concerned with asymptotic or minimax performance: how well does the algorithm perform in large training sets assuming nature is adversarial. In other words, how well does an algorithm perform in the worst-case; the true generating distribution has been selected to trick the algorithm.

[1] A parametric model will summarize what it learns from the data in the parameters. The standard example is linear regression: $y_i = a x_i + b$. The parameters are $\theta = \{a, b\}$. There are also nonparametric models where the training data cannot be completely summarized in a finite number of parameters. We will use the term parameters in both cases. There are also hyper-parameters, which are the parameters of probability distributions placed on parameters.

ISBN 978-0-9566139-1-2

Bayesian methods always provide a predictive distribution p while it is common for non-Bayesian methods to just specify a single best prediction a. When a point estimate is required (1) must be used to convert the predictive distribution p to an action a.

CHANGE POINT MODELS

The BOCPD algorithm introduced in [1] efficiently finds a distribution on the time since the last change point, namely the *run length*. BOCPD requires one to specify an underlying predictive model (UPM) and a hazard function. The UPM is used to compute $p(x_t|x_{t-\tau:t},\theta_m)$ (known as a posterior predictive) for any $\tau \in [1,\dots,(t-1)]$, at time t. Intuitively, the UPM is a simple model of the time series x that changes at every change point. The hazard function $H(\tau|\theta_h)$ describes how likely a change point is given the current run length τ. We define the change point vector $c_t \in \{0,1\}$ to be true (1) if there is a change point at time t and false (0) otherwise. Learning the hyper-parameters and improving computational efficiency have been addressed in [13].

Example BOCPD Model

A simple example of BOCPD would be to use a constant hazard function $H(\tau \mid \theta_h) := \theta_h$, meaning $p(\tau_t = 0 \mid \tau_{t-1}, \theta_h)$ is independent of τ_{t-1} and is constant, giving rise to geometric inter-arrival times for change points. A constant hazard implies the presence of a change point at time t is determined by a weighted coin flip independent of how long it has been since the last change point. The geometric inter-arrival distribution is the discrete time analog of the exponential distribution, which is often used to model the arrival of buses or radioactive decay where events are unrelated to one another. We can use a UPM with iid Gaussian observations with changing mean and variance; we apply a distribution known as the Normal-Inverse-Gamma on the mean and variance, which is computationally advantageous:

$$x_t \sim N(\mu, \sigma^2), \qquad \mu \sim N\left(\mu_0, \frac{\sigma^2}{\kappa}\right), \qquad \sigma^{-2} \sim gamma(\alpha, \beta), \tag{2}$$

here the model hyper-parameters are $\theta_m := \{\mu_0, \kappa, \alpha, \beta\}$. A sample draw from this model is shown in Figure 1. Sampling data from a model can be a good way to visualize the assumptions it makes.

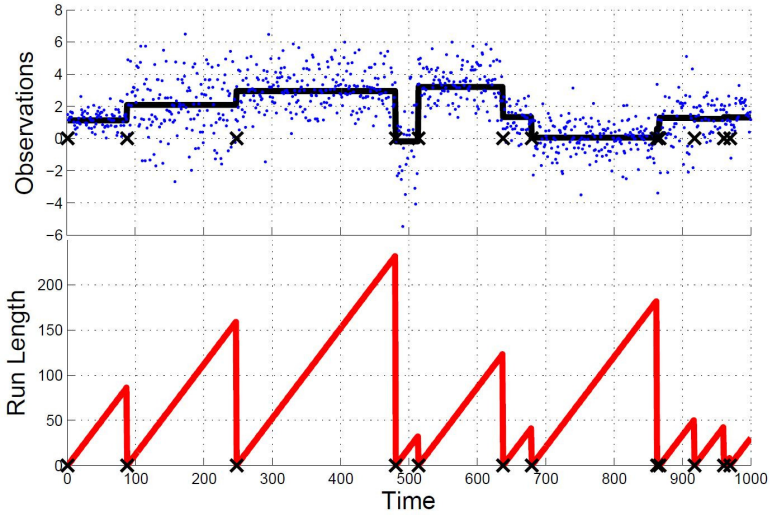

Figure 1: Sample draw from BOCPD with a Gaussian UPM and constant hazard. Top panel: The blue dots are the data points x_t ; at each time point t, x_t is sampled from (2a). The solid black line represents the mean in each regime. Note that the mean and variance is changing in each regime; at each change point the mean and variance is a new sample from (2b). Bottom panel: The black crosses are the change points and the solid red line is the run length.

ISBN 978-0-9566139-1-2

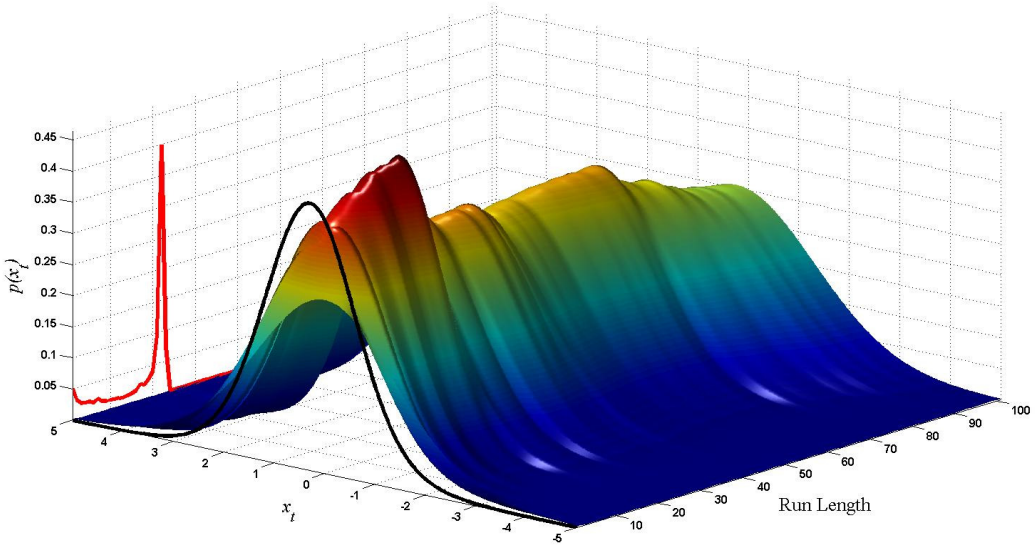

Figure 2: Demonstration of BOCPD on a sample draw with a Gaussian UPM and constant hazard. The surface plot is the predictive distribution $p(x_{t+1}, r_t | x_{1:t})$ conditional on the run length r_t at each time step using the UPM as in (5). The solid red line is the run length distribution $p(r_t | x_{1:t})$ from (4), while the solid black line is the marginal distribution $p(x_{t+1} | x_{1:t})$ using (3).

The BOCPD Algorithm

This section describes the core of the BOCPD method and thus assumes a level of comfort using the manipulations of probability theory. BOCPD calculates the run length predictive distribution at time t, i.e. $p(r_t | x_{1:t})$, in a recursive fashion. At every time step t we use the last run length distribution $p(r_{t-1} | x_{1:t-1})$ to efficiently calculate $p(r_t | x_{1:t})$. We can make predictions about the next data point x_t in a way that maintains our uncertainty in the true run length, through marginalization of the run length variable:

$$p(x_{t+1} | x_{1:t}) \tag{3}$$
$$= \sum_{r_t} p(x_{t+1} | x_{1:t}, r_t)\, p(r_t | x_{1:t})$$
$$= \sum_{r_t} p(x_{t+1} | x_t^{(r)})\, p(r_t | x_{1:t})$$

where $x_t^{(r)}$ refers to the last r_t observations of x, and $p(x_{t+1} | x_t^{(r)})$ is computed using the UPM. The run length distribution can be found by normalizing the joint likelihood:

$$p(r_t | x_{1:t}) = \frac{p(r_t, x_{1:t})}{\sum_{r_t} p(r_t, x_{1:t})}. \tag{4}$$

ISBN 978-0-9566139-1-2

The joint likelihood can be updated online recursively:

$$\gamma_t := p(r_t, x_{1:t}) = \sum_{r_{t-1}} p(r_t, r_{t-1}, x_{1:t})$$
$$= \sum_{r_{t-1}} p(r_t, x_t \mid r_{t-1}, x_{1:t-1})\, p(r_{t-1}, x_{1:t-1}) \tag{5}$$
$$= \sum_{r_{t-1}} \underbrace{p(r_t \mid r_{t-1})}_{\text{hazard}} \underbrace{p(x_t \mid r_{t-1}, x_t^{(r)})}_{\text{UPM}} \underbrace{p(r_{t-1}, x_{1:t-1})}_{\gamma_{t-1}}$$

This defines a forward *message passing* [9] scheme to recursively calculate γ_t from γ_{t-1}. The conditional can be restated in terms of messages as $p(r_t \mid x_{1:t}) \propto \gamma_t$. The calculations in this section are illustrated visually in Figure 2. All the distributions mentioned so far are implicitly conditioned on the set of hyper-parameters ϑ.

The example in Section 4.1 used a UPM that assumed the data was iid in each regime. Gaussian processes (GPs) [10] are probability distributions over generic functions. BOCPD has been combined with Gaussian process time series models in [11]. While [13] extended BOCPD to learn the hyper-parameters θ using the time series $x_{1:T}$ only (unsupervised), [12] incorporates labeled data for when the change points occur $c_{1:T}$ to better learn θ.

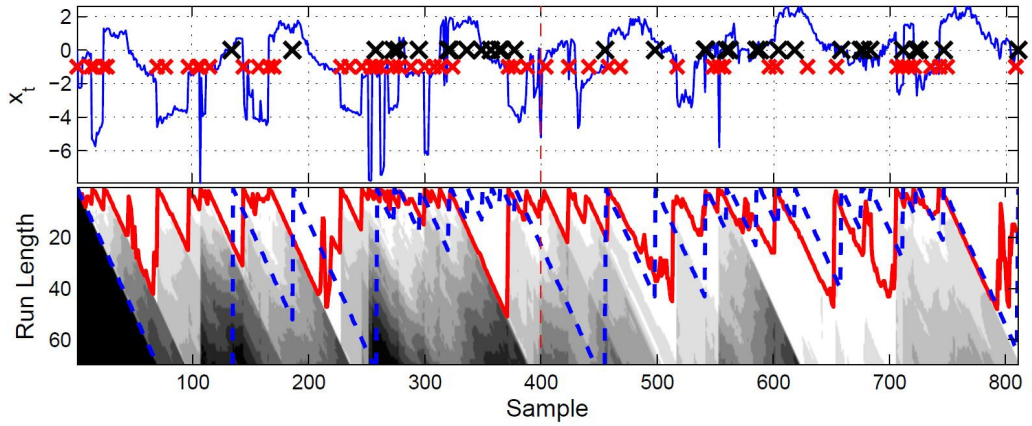

Figure 3: Modem Data: the top panel shows the modem transmission strength in blue. The labeled change points, the power supply alarm, are shown as black crosses. The red crosses represent alerts, which are placed when the probability of a change point since the last alert exceeds 95%. The bottom panel shows the log run length distribution (QuickTime™ and a decompressor are needed to see this picture.) **(lighter means higher probability). The red line indicates the median of the distribution while the dashed blue line is the labeled run length. Everything before the vertical dashed line is used to train the hyper-parameters** .

ISBN 978-0-9566139-1-2

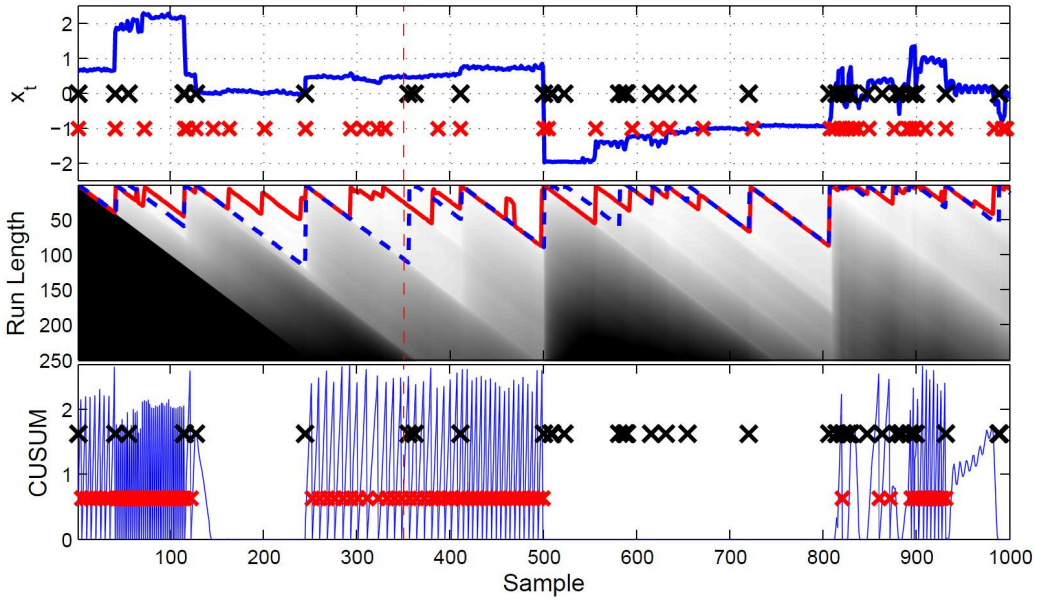

Figure 4: Amplifier Data: the top panel shows the amplifier power in blue. The labeled change points are shown as black crosses. The graphical conventions are the same as those in Figure 3 The bottom panel shows the output of CUSUM. The red crosses are the CUSUM alert and the blue line is the CUSUM test statistic , which was reset to zero at each alert.

SATELLITE DATA

In this section we look at fault data from a satellite base station provided by a major satellite communications corporation. We use faults or alarms as change points and sensor measurements as when training BOCPD. In this context, the run length is the time since the last fault, which might be useful for faults that are not always easily detectable. We can also provide a distribution on time until next change points as in a prediction on time until the next fault using the hazard function learned in training. We consider both unsupervised, which ignores the known change points in training [13], and supervised, uses the change points in training [12], to train the hyper-parameters series that we analyze

Table 2 The results are provided with 95% error bars and the p-value testing the null hypothesis that methods are equivalent to the best performing method, according to NLL, using a one sided t-test. The first three columns represent the loss in predicting the run length (using) while the final column shows the loss in data space of predicting the next data point (using). The units of NLL are nats/observation. For continuous variables (data space prediction) we shift the NLL so the best method has NLL zero. The MAE and MSE have the units of and , respectively, in run length prediction. Variants of BOCPD are marked with *.

Method	NLL x 10^1	p-value	MAE x 10^0	p-value	MSE x 10^-2	p-value	NLL x 10^1	p-value
Modem Data (350 Training Points, 460 Test Points)								
CUSUM	46.45±0.24	<0.0001	28.8±1.3	<0.0001	11.85±0.78	<0.0001	N/A	N/A
TIM	41.30±0.49	0.069	15.50±0.92	0.1417	5.30±0.37	<0.0001	7.24±0.87	<0.0001
Unsupervised *	133±15	<0.0001	18.3±1.7	0.0002	6.22±0.97	0.0002	0.1±1.3	0.4617
Supervised *	**40.74 ± 0.55**	N/A	**14.77 ± 0.97**	N/A	**4.33 ± 0.35**	N/A	**0.0 ± 1.4**	N/A
Amplifier Data (350 Training Points, 650 Test Points)								
CUSUM	44.28±0.26	<0.0001	22.57±0.95	<0.0001	8.48±0.54	<0.0001	N/A	N/A
TIM	44.28±0.37	<0.0001	21.79±0.94	<0.0001	11.02±0.65	<0.0001	37.7±1.4	<0.0001
Unsupervised *	103±16	<0.0001	11.4±1.4	0.0183	3.97±0.77	0.0032	**0.0 ± 1.4**	0.953
Supervised *	**37.0 ± 1.1**	N/A	**9.6 ± 1.1**	N/A	**2.66 ± 0.56**	N/A	1.5±1.2	N/A

ISBN 978-0-9566139-1-2

In the first example, we use alarms from a power supply as the labeled change points c and a modem transmission strength as the time series x. The variables share only a tenuous link and the two are not obviously related upon visual inspection. However, supervised BOCPD still manages to find a signal. The time series was differenced and standardized prior to training. We show the run length distribution of BOCPD and the time series in Figure 3.

We trained on the first 350 time points and tested on the remaining 460 points in Table 2; in all of the experiments we compare our results to the time independent model (TIM) as a reference. In data space prediction the TIM model corresponds to modeling the data as iid Gaussian. In the run length space the TIM model corresponds to assuming a constant hazard and an uninformative UPM $p(x_t) \propto 1$, which gives a geometric run length distribution in large t. We also compared to the CUSUM method [8], which is a classic frequentist change point detection method. The CUSUM maintains a test statistic S_t, which one uses to alert when it crosses a threshold. We set the parameters of the CUSUM as recommended by [7]. The threshold is chosen to control the false positive rate. However, this does not provide predictions needed for evaluation. Consequently, the output of CUSUM in training was fit to the labeled run length using a linear model.

We also consider a time series of amplifier power. The correspondence between the change points and the time series is much closer in this time series than in the modem transmission strength. We see in Figure 4 that the median of the run length distribution almost perfectly matches the true run length. We also show the output of CUSUM on the same time series. It alerts far too often, it seems to have trouble with the recurrent nature of the change points in this problem. Note that the CUSUM is guaranteed not to alert more than 5% of the time if there is no change point. However, it does not say that it will find where the change points are in the event of recurring change points. We see that this is a problem for CUSUM in Figure 4. Quantitative results are shown in Table 2.

CONCLUSIONS

Change point detection is a widely applicable machine learning technique and we have shown it has many uses when applied to monitoring electronic systems. BOCPD presents a general method that allows for many other simpler models to be used as UPMs. It is therefore a very important part of time series analysis and machine learning. We trained and evaluated BOCPD from the machine learning perspective, we set the hyper-parameters according to training data and checked its predictive performance on a separate test set. We have shown that just controlling a false positive rate is insufficient for good predictive performance in complex change point problems, such as satellite fault logs. There is much room for extensions within the framework provided by BOCPD.

REFERENCES

[1] R. P. Adams and D. J. C. MacKay. Bayesian online changepoint detection. Technical report, University of Cambridge, Cambridge, UK, 2007.
[2] D. Barry and J. A. Hartigan. A bayesian analysis of change point problems. *Journal of the American Statistical Association*, 88:309–319, 1993.
[3] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2007.
[4] A. P. Dawid. Probability forecasting. *Encyclopedia of Statistical Sciences*, 7:210–218, 1986.
[5] A. P. Dawid. The geometry of proper scoring rules. Technical Report 268, Department of Statistical Science, University College London, April 2006.
[6] P. Fearnhead and Z. Liu. Online inference for multiple changepoint problems. *Journal of the Royal Statistical Society*, 69:589–605, 2007.
[7] Olivia A Grigg and David J Spiegelhalter. An empirical approximation to the null unbounded steady-state distribution of the cumulative sum statistic. *Technometrics*, 4:501–511, November 2008.

ISBN 978-0-9566139-1-2

[8] E. S. Page. A test for a change in a parameter occurring at an unknown point. *Biometrika*, 523–527:42, 1955.

[9] Judea Pearl. Reverend Bayes on inference engines: A distributed hierarchical approach. In *AAAI*, pages 133—136, Pittsburgh, PA, USA, August 1982.

[10] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

[11] Yunus Saatci, Ryan Turner, and Carl Edward Rasmussen. Gaussian process change point models. In *Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel, June 2010.

[12] Ryan Turner and Carl Edward Rasmussen. Supervised Bayesian online change point detection, 2010. In Press.

[13] Ryan Turner, Yunus Saatci, and Carl Edward Rasmussen. Adaptive sequential Bayesian change point detection. In Zaid Harchaoui, editor, *Temporal Segmentation Workshop at NIPS 2009*, Whistler, BC, Canada, December 2009.

[14] X. Xuan and K. Murphy. Modeling changing dependency structure in multivariate time series. In *Proceedings of the 24th International Conference on Machine Learning*, pages 1055–1062, Corvallis, OR, USA, June 2007.

DIVERSE ENGAGEMENT: DRAWING IN THE MARGINS,
Proceedings of the Interdisciplinary Graduate Conference, Cambridge University, UK (28-29, June 2010)
ISBN 978-0-9566139-1-2 © University of Cambridge, Graduate Development Programme.

IMPACT ASSESSMENT TOOLS IN PLANNING FOR SUSTAINABLE DEVELOPMENT OF CULTURALLY SIGNIFICANT URBAN AREAS

Tatiana Vadimovna VAKHITOVA

PhD Student[1], University of Cambridge, UK
tvv21@cam.ac.uk

Keywords: Cultural Heritage, Values, Impact Assessment, Sustainable Development

ABSTRACT

This background paper describes research in progress and develops two major topics: the role of cultural heritage in sustainable development of cities and the challenges of considering cultural heritage in planning impact assessment tools. The article will conclude with a section on the need for further research.

INTRODUCTION

Cultural heritage has diverse values and therefore necessitates the participation of different stakeholders, such as heritage experts, city planners, architects, and local community groups in the urban planning process. The PhD research project considers the assumption that impact assessment tools, such as environmental or social impact assessments, which are used to facilitate decision-makers in their choice of a more sustainable alternative, provide a useful participation-based framework and finds that the way diverse heritage values are introduced in assessment practice is not adequate. The project focuses on this problem by analyzing the process of identification and communication of cultural heritage values in the example of World Heritage sites in the UK with officially identified Outstanding Universal Value.

The role of Cultural Heritage in cities' Sustainable Development

Built cultural heritage plays an important role in the economic and social fabric of a city, in addition to its direct contribution to the bio-physical environment. The historic environment adds to national and local community identity, which is important for sense of place and social cohesion. Urban regeneration should therefore recognize local values, utilize local knowledge and expertise, and provide a sense of ownership to stakeholders, such as residents and other beneficiaries in the area; as Evans ascertains '*culture is a critical aspect of mediating and articulating community need, as development is planned and takes shape, through culture's potential to empower and animate*'[2] (see also Folke[3]; Raymond *et al.*[4]).

The role of the historic built environment in promoting economic growth is also acknowledged. Heritage can boost the local and national economy and create jobs by attracting tourists and investment, providing leisure, recreation, and educational facilities (see e.g. PPS 15[5]; Tweed[6], Nijkamp and Riganti[7]; EH Guidance Note to Circular[8]). Moreover its '*indirect spin-offs can benefit the construction and service industries*'[9], i.e. adding value to the area/property. Therefore, cultural heritage must be seen as '*part of the larger sphere of socio-cultural processes*'[10] and managed in such a way as to '*be able to generate real economic and social benefits for their local host communities*'[11].

Many authors agree that cultural heritage should be promoted as a vehicle for social and economic regeneration if we are to progress towards sustainable development (see e.g. Rodwell[12]; Lazrak *et al.*[13]; Nijkamp and Riganti[14]; Evans[15]; Stubbs[16]; Jones and Slinn[17]). Lazrak *et al.* notes that cultural heritage is a 'critical infrastructure' of

ISBN 978-0-9566139-1-2

modern cities, which provides '*long-lasting and sustainable anchor points that may create stability in a permanently fluctuating and competitive environment*'[18]. The importance of cultural, artistic and innovative sectors for urban vitality and the creation of wealth is witnessed in many scholars' works. As stated by Catterall: '*The task is to develop an understanding (including methods of study) of the ways—cultural and ethical—in which even the 'worst estates' can take part in and help shape the relics of their city (and society) as well as their locality*'[19].

A recent issue of the Built Environment Journal *Creative Knowledge City* (Vol. 35 №2, 2009) stresses the importance of both physical surroundings and the socio-cultural dimensions of cities to attract and retain qualified workers in creative industries. The factors identified as contributing to the overall quality of place, and therefore influencing peoples' choice of location, include diversity, tolerance and safety, environmental quality, aesthetics, amenities, opportunities for recreation, culture and an environment supportive of lifestyle choices. Brown and Meczynsky[20] state that 'multifaceted' cities have to address these issues if they want to be competitive in the 'knowledge-driven' economy. The role of cultural heritage should not be underestimated.

Among the aspects which characterize the quality of place, as further in Brown and Meczynski, authenticity and livability are particularly hard to construct or plan; therefore it is important to sustain or '*create favorable conditions for them to develop*'[21]. The argument put forward is that a '*more likely means of success is to build on what is already there, enhancing the distinctive and unique qualities of the existing urban environment*' (*Ibid.*) rather than the 'iconic' architecture used as a flagship to attract tourists. Van Oers reinforces the statement: '*more and more cities are pushed into the role of drivers of regional growth and development*'[22] and by trying to capture capital, historic cities have *'leverage in being able to offer something unique*' (*Ibid.*).

There is an extensive body of literature on cultural economics, and specifically on economic methods for assessing cultural heritage values. Remarkable is the interest in connections between decisions about economic development and the cultural aspects of cities. Mason argues that '*if planned carefully, cities can thrive economically and culturally, and these two types of value are closely connected*'[23]. Mason further states that historic built environments and the values attached to them '*comprise not just an amenity but an infrastructure supporting regeneration by providing flows of both cultural value and economic value'* (*Ibid.*, see also Throsby[24]), which should be therefore regarded as opportunities and considered both as cultural and economic assets.

Cultural Heritage in Impact Assessment Tools

Estimation of the social merits of cultural assets is challenging, and a better understanding of how people interact with their urban environment is required (see e.g. Tweed[25]; Nijkamp and Riganti[26]). Transformation of the built environment through new development has an immediate positive or negative impact on the local population and therefore can potentially influence the overall results of projects (e.g. SUIT[27]). Nijkamp and Riganti argue that '*The public's preferences for aesthetic and use attributes are rarely elicited, despite their potential importance in decision-making*'[28]. Raymond supports this by saying that '*whilst biophysical, and increasingly economic, values are often used to define high priority hotspots in planning for conservation and environmental management, community values are rarely considered*'[29]. Moreover as Tweed states '*culture is not generally recognized in urban policy or environmental and quality of life indicators (such as health, education, employment, crime) and therefore is absent from regeneration measurement criteria*'[30].

Nijkamp and Riganti[31] recognize that research efforts have insufficiently integrated to tackle the complex issues related to heritage conservation. Therefore, as the authors suggest, there is a need to develop comprehensive approaches and methodologies for adequate consideration of cultural heritage in management and planning for cities sustainable development. The argument put forward is that valuation methods and assessment procedures should help to '*better integrate conservation in the social agenda, enhancing social justice and equity in the provision and management of cultural heritage and therefore, play a part in the assessment of progress towards city sustainable development*'[32]. For successful integration, 'culture' requires planning, investment, infrastructure, development and design decisions and it must not be considered in isolation from other urban planning decisions. Moreover, as Mason advocates, '*conservation of the historic built environment (not just buildings, but landscapes and larger districts, too) is a key function of healthy cities*'[33]. Therefore, conservation should be an important part of urban planning policy and design.

The European Strategic Environmental Assessment (SEA) and Environmental Impact Assessment (EIA) Directives both require consideration of cultural heritage although in practice its performance is still relatively weak. Different values cultural heritage has are usually only very narrowly presented in the assessments (e.g.

ISBN 978-0-9566139-1-2

Jones and Slinn[34]). Better consideration of cultural heritage values, through assessment and appropriate actions, can help to minimize damage or loss caused by development projects. It is claimed in SUIT that the framework, imposed by EIA and SEA procedures, is important to use in '*sensitive decision-making processes in relation to cultural heritage values'* [35] as it is '*likely to help in establishing a constructive debate or dialogue between all the concerned actors*' (*Ibid.*). Bond and Teller also encourage a more systematic application of the SEA procedure to the local plans '*directly affecting sensitive urban historical areas'*[36]. Furthermore, as in Bond *et al.* [37], the assessment procedures need to take place within a general framework of a decision-making process where consensus at one level must be reached before progressing to the next.

Literature review has revealed several challenges for better introduction of cultural heritage in Impact Assessment Tools: difficulties in identification and communication of cultural heritage values, need for wider and earlier stakeholder engagement (community and experts), lack of specific guidance, and the knowledge of assessment processes among cultural experts and vice-versa. This problem of communication deficiencies leads to connected inadequate identification of cultural values and their explicitness. There is a need for an introduction of more qualitative indicators, which better reflect intangible vales cultural heritage possess. And of general importance, post-monitoring/evaluation procedures have to be developed in a participative way throughout the assessment process.

CONCLUSION

The research project is designed to respond to identified problem field. The ultimate aim of the research is to facilitate the planning decision-making process in culturally significant urban areas, based on the values and knowledge of the local community, by encouraging informed participation of key stakeholders. The research sets out, firstly, to, provide evidence for structuring of cultural heritage values in a framework for use by planners and developers; secondly, to develop recommendations on the design of the impact assessment process for culturally significant urban areas; and thirdly, to identify international cases of best practice as a valuable benchmark for both city planners and developers to use in strategic planning for conservation and development.

The methodology includes both quantitative and qualitative approaches: case studies and surveys' analyses. Data collected through interviews and through desk-based research will be compared with those collected through surveys. Part of the research methodology includes several seminars with focus groups from academia and practice to collect their feedback on initial research assumptions.

Chosen case studies comprise World Heritage (WH) sites in the UK. In particular, the research will focus on sites over which UNESCO has expressed concern regarding existing/potential negative impacts of developments on Outstanding Universal Value (OUV). Statistics shows that in 2007 40% of WH sites reported potential negative impacts of urban development and regeneration projects on OUV, as presented by Van Oers[38]. The research will add to understanding of the reasons underlying this negative trend.

Understanding the context (e.g. planning practice, organizational structure, norms and values), where the Impact Assessment Tools are applied, should reveal barriers and opportunities for a better communication process among the key stake holders, regarding cultural heritage values. It will also help to uncover the potential for improvements in their efficiency. The critical role of the context, including the '*extent of the political commitment to sustainable development*' (Wallington *et al.*[39]) and how the impact assessment process is fitted into planning process itself (see e.g. Stoeglehner *et al.*[40]) have been identified by many scholars.

The PhD research also contributes to an international research project *Outstanding Universal Value of World Heritage Cities and Sustainability*. This is a collaboration work with UNESCO WH Centre and Eindhoven University of Technology in Holland. The findings will be used in a web based tool, which maps the OUV assessment process, and thus leads to an increased transparency and sustainability in planning decision-making practices of WH sites worldwide.

ISBN 978-0-9566139-1-2

REFERENCES

[1] **Note: Acknowledgments**
Dr Peter Guthrie (University of Cambridge), Dr Theophilus Hacking (University of Cambridge), Dr Minna Sunikka-Blank (University of Cambridge) and Dr Ana Pereira Roders (TU/e University).

[2] Evans, G. *Measure for Measure: Evaluating the Evidence of Culture's Contribution to Regeneration* (2005) Urban Studies, 42, pp. 959-983, p. 959.
[3] Folke, C. *Resilience: the emergence of a perspective for social–ecological systems analyses.* (2006) Global Environmental Change 16, pp. 253-267.
[4] Raymond, C.M. *et al. Mapping community values for natural capital and ecosystem services* (2009) Ecological Economics, 68, pp. 1301-1315.
[5] English Heritage, *PPS 15: Planning for the Historic Environment: Historic Environment Planning Practice Guide* (2009) from http://www.english-heritage.org.uk.
[6] Tweed, C., Sutherland, M. *Built cultural heritage and sustainable urban development* (2007) Landscape and Urban Planning, vol. 83 (1), pp. 62-69.
[7] Nijkamp, P. and Riganti, P. *Assessing cultural heritage benefits for urban sustainable development* (2008) Int. J. Services Technology and Management, Vol. 10, (1), pp. 29-37.
[8] English Heritage *Guidance Note to Circular for England on the Protection of World Heritage Sites* (2009) from http://www.english-heritage.org.uk.
[9] IAIA *Cultural Heritage in Environmental Assessment* (1994) The World Bank Number 8, ch. 3, p.2.
[10] Avrami, E., de la Torre, M., Mason, R. (eds.) *The Values and Heritage Conservation: Research Report* (2000) Los Angeles: Getty Conservation Institute, from www.getty.edu/conservation/ p. 11.
[11] Hampton, M. *Heritage, Local Communities and Economic Development* (2005) Annals of Tourism Research, vol.32 (3), pp. 754.
[12] Rodwell, D. *Finding the Spirit of Place* (2008) delivered at the 16th ICOMOS General Assembly, 1 October, Workshop: Historic Urban Landscapes: Concept and Management, Quebec.
[13] Lazrak, F., Nijkamp, P., Rietveld, P. and Rouwendal, J. *Cultural Heritage and Creative Cities: an Economic Evaluation Perspective* (2009) (working paper) Serie Research Memoranda 2009-30, VU University Faculty of Economics, Amsterdam.
[14] Nijkamp, P. and Riganti, P. *Assessing cultural heritage benefits for urban sustainable development* (2008) Int. J. Services Technology and Management, Vol. 10, (1), pp. 29-37.
[15] Evans, G. *Measure for Measure: Evaluating the Evidence of Culture's Contribution to Regeneration* (2005) Urban Studies, 42, pp. 959-983.
[16] Stubbs, M. *Heritage-Sustainability: Developing a Methodology for the Sustainable Appraisal of the Historic Environment* (2004) Planning, Practice & Research, Vol. 19, (3), pp. 285-305.
[17] Jones C.E. and Slinn P. *Cultural Heritage in EIA – Reflections on Practice in North West Europe* (2008) Journal of Environmental Assessment Policy and Management, V. 10, 3, pp. 215-238.
[18] Lazrak, F., Nijkamp, P., Rietveld, P. and Rouwendal, J. *Cultural Heritage and Creative Cities: an Economic Evaluation Perspective* (2009) (working paper) Serie Research Memoranda 2009-30, VU University Faculty of Economics, Amsterdam, 20 pages, pp. 2-3.
[19] Catterall, B. *Culture as a critical focus for effective urban regeneration* (1998) Town & Country Planning Summer School, University of York (in Evans: 2005, 978): Evans, G. *Measure for Measure: Evaluating the Evidence of Culture's Contribution to Regeneration* (2005) Urban Studies, 42, pp. 959-983.
[20] Brown, J., Meczynski, M. *Complexcities': Locational Choices of Creative Knowledge Workers* (2009) in Built Environment, *Can We Plan the Creative Knowledge City?*, vol.35 (2) (Guest eds. Caroline Chapain, Chris Collinge, Peter Lee and Sako Musterd), Alexandrine Press ed. Peter Hall, David Banister, pp. 238-252.
[21] *Ibid.* 250.
[22] Van Oers R. *Towards new international guidelines for the conservation of historic urban landscapes (HUL)s* (2007) City & Time, 3 (3), p. 46.
[23] Mason, R. *Culture and Historic Preservation in Urban Revitalization: Report from North America* (2009) (chapter from Visby book in press), pp. 100-117, p. 101.
[24] Throsby, D. *Economics and Culture* (2001), New York: Cambridge University Press.
[25] Tweed, C., Sutherland, M. *Built cultural heritage and sustainable urban development* (2007) Landscape and Urban Planning, vol. 83 (1), pp. 62-69.
[26] Nijkamp, P. and Riganti, P. *Assessing cultural heritage benefits for urban sustainable development* (2008) Int. J. Services Technology and Management, Vol. 10, (1), pp. 29-37.
[27] SUIT Report *Sustainable development of Urban historic areas through an active Integration within Towns* (2004) European Commission, Directorate-General for Research Information and Communication Unit, Sector – Urban sustainability and cultural heritage, Research Report # 16, Project n° EVK4-2000-00017, 128 pages.
[28] Nijkamp, P. and Riganti, P. *Assessing cultural heritage benefits for urban sustainable development* (2008) Int. J. Services Technology and Management, Vol. 10, (1), pp. 29-37, p.36.

ISBN 978-0-9566139-1-2

[29] Raymond, C., Bryan, BA, MacDonald AC, *et al. Mapping community values for natural capital and ecosystem services* (2009) Ecological Economics, 68 , pp. 1301-1315, p.1301.
[30] Tweed, C. and Sutherland, M. *Built cultural heritage and sustainable urban development* (2007) Landscape and Urban Planning, vol. 83 (1), pp. 62-69, p.62.
[31] Nijkamp, P. and Riganti, P. *Assessing cultural heritage benefits for urban sustainable development* (2008) Int. J. Services Technology and Management, Vol. 10, (1), pp. 29-37.
[32] *Ibid.* 35.
[33] Mason, R. *Culture and Historic Preservation in Urban Revitalization: Report from North America* (2009) (chapter from Visby book in press), pp. 100-117, p. 102.
[34] Jones, C.E. and Slinn, P. *Cultural Heritage in EIA – Reflections on Practice in North West Europe* (2008) Journal of Environmental Assessment Policy and Management, V. 10, 3, pp.215-238.
[35] SUIT Report *Sustainable development of Urban historic areas through an active Integration within Towns* (2004) European Commission, Directorate-General for Research Information and Communication Unit, Sector – Urban sustainability and cultural heritage, Research Report # 16, Project n° EVK4-2000-00017, p.22.
[36]Bond, A., Teller., J. *The Consideration of Cultural Heritage in present European Environmental Policies* (2004) 101-102, In SUIT Report *Sustainable development of Urban historic areas through an active Integration within Towns* (2004) European Commission, Directorate-General for Research Information and Communication Unit, Sector – Urban sustainability and cultural heritage, Research Report # 16, Project n° EVK4-2000-00017, p. 102.
[37] Bond, A., Langstaff, L., Ruelle, C. *Monitoring and Post-Evaluation of the Cultural Heritage Component of Environmental Assessments* (2004) pp. 111-113, In SUIT Report *Sustainable development of Urban historic areas through an active Integration within Towns* (2004) European Commission, Directorate-General for Research Information and Communication Unit, Sector – Urban sustainability and cultural heritage, Research Report # 16, Project n° EVK4-2000-00017.
[38] Van Oers R. *Towards new international guidelines for the conservation of historic urban landscapes (HUL)s* (2007) City & Time, 3 (3), p. 46.
[39] Wallington, T., Bina, O., Thissen, W. *Theorising strategic environmental assessment: Fresh perspectives and future challenges* (2007) Environmental Impact Assessment Review, 27, pp. 569–584.
[40] Stoeglehner, G., Brown, A.L., Kørnøv, Lone B. *SEA and planning: 'ownership' of strategic environmental assessment by the planners is the key to its effectiveness* (2009) Impact Assessment and Project Appraisal, 27 (2), pp. 111-120.