Citation for published version


DOI

https://doi.org/10.1109/ISTAS.2015.7439436

Link to record in KAR

https://kar.kent.ac.uk/58171/

Document Version

Pre-print

Copyright & reuse
Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research
The version in the Kent Academic Repository may differ from the final published version. Users are advised to check http://kar.kent.ac.uk for the status of the paper. Users should always cite the published version of record.

Enquiries
For any further enquiries regarding the licence status of this document, please contact: researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at http://kar.kent.ac.uk/contact.html
New Technologies and Mixed-Use Convergence
How Humans and Algorithms are Adapting to Each Other

Sally A. Applin/Michael D. Fischer
University of Kent, Canterbury, School of Anthropology and Conservation
Centre for Social Anthropology and Computing
Canterbury, UK
sally@sally.com/m.d.fischer@kent.ac.uk

Abstract—Human experience with technology has shifted from technological contexts occasionally requiring intervention by a fraction of people mostly in command of technologies, to technological contexts that require constant ongoing participation from most people to complete tasks. We examine the current state of 'mixed-use' new technologies integration with legacy systems, and whether the human assistance required to complete tasks and processes could function as a training ground for future smart systems, or whether increasing 'co-dependence with' or 'training of' algorithmic systems, enhancing task completion and inadvertently educating systems in human behaviour and intelligence, will simply subsume people into the algorithmic landscape.

As the Internet of Things (IoT) arises in conjunction with advancing robotics and drone technology, semi and fully automated algorithmic systems are being developed that intersect with human experience in new and heterogeneous ways. Many new technologies are not yet flexible enough to support the choices people require in their daily lives, due to limitations in the algorithmic 'logics' used that restrict options to predetermined pathways conceived of by programmers. This greatly limits human agency, and presently the potential to overcome problems that arise in processes. In this mixed-use period, we have the opportunity to develop new ways to address ethical guidance as knowledge that machines can learn. We explore promoting embedding of ethically-based principles into automated contexts through: 1) developing mutually agreed automated external ethical review systems (human or otherwise) that evaluate conformance across multiple ethical codes and provide feedback to designers, agents, and users on the distribution of conformance; 2) focussing on review systems to drive distributed development of embedded ethical principles in individual services by responding to this feedback to develop ongoing correction through dynamic adaption or incremental releases; and 3) using multi-agent simulation tools to forecast scenarios in real time.

Keywords—automation; agency; algorithms; anthropology; ethics; knowledge

I. INTRODUCTION

Many new technological systems (and technological agents) are unprecedented with respect to capabilities to quantify and track people and themselves, and to do so while disrupting established systems. As a new technology 'disrupts' legacy industries and institutions, people are expected to rapidly adjust to a new and increasingly heterogeneous technological presence. Systems that have seemed relatively stable for some time (governance, postal services, transportation, agriculture, etc.) are disrupted by new technologies, often funnelled through mobile devices or distributed though algorithms in ubiquitous machines such as ticket vending machines, cash registers, airport computers, parking lot machines, laundry machines etc. Many times the algorithms for these machines are more rigid than situated use cases require and people must find workarounds to achieve an outcome [1; 2], modifying their behaviour and the context of use so that a machine's programming can address what they are trying to do. We refer to this behaviour as an outcome of covert agency [3].

Covert agency addresses some of the questions arising from discussions of "workarounds" such as those by Woods and Hollnagel (2006) [1] and Woods et al. (1994) [2]. In many cases employing covert agency to engineer a workaround is a "critical" element in the successful achievement of an outcome. Disguising these "deviant" choices is also critical, as these choices nominally subvert workplace management, scripts and processes and/or accepted use of automated equipment in the work process, however inadequate these might be. In short, 'hacking' becomes a critical workplace skill as inflexible systems and algorithms in the workplace become more and more commonplace. The transition to the IoT, with overt distribution of function, will require these skills, perhaps finally removing the need for covertness.

While covert behaviours are not tracked and recorded due to being cloaked, behaviours associated with making a sanctioned (e.g. scripted, restricted) choice are recorded. This information has become
valuable and used to improve the design of systems whose developers want to improve their programs' ability to interact, offer and effect choices that meet human and institutional needs.

We examine the current shift towards automated processes and services and the way that humans and algorithms are adapting to each other during this transition. We also explore how a variety of ethically-based frameworks might be integrated across automated systems via agent awareness and brokering.

II. THE MIXED-USE WORLD

A. An Adjustment Period

In this mixed-use world of new and legacy systems, humans and machines are in an adjustment period [1]. The fast rate of automation innovation associated with all aspects of living seems at times to outpace how fast we are able to adjust. How we regulate new technologies and develop sensible usage cases become new issues for us to resolve. Furthermore, developers of new technology often focus on single individual use cases (the single person and their drone, the car and the road, a worker and machine) leaving aside the impact of how the heterogeneous requirements of all of these new technologies will co-exist in the hands of both the people using a technology and the people it is disrupting. Although many people deal with this situation by simply ignoring technologies until they can deal with these, foregoing possible new capabilities and/or benefits, others simply refuse to engage with them.

Concurrently, the script, process and machine automation-driven world humans are creating is adjusting to human usage and interaction [3]. Some machines record actions for improvement to their process but others do not engage with needs outside of their programmed script for how a transaction should be executed. The opportunities for people to avoid automation and make choices that are less technologically pre-determined will become more rare, particularly in cases where the automation process is not designed to learn from human interaction. This will be particularly critical with the Internet of Things (IoT) yielding near ubiquitous systems intersecting nearly all aspects of our lives.

B. Automated Financial Technologies

Automated financial technologies require consumers' trust at an end-user level. For example, Automated Teller Machines (ATM's) (also known as Cash point or Cash machines in the UK) are a technology that was not fully welcomed by everyone when initially released as an alternative to human tellers for banking activities [4]. People mistrusted early ATMs' ability to count bills and were concerned that they would have no recourse for machine errors [5]. As such, ATMs never replaced banking staff entirely. Gradually, ATM functions were embedded into other contexts offering "cash back," such as grocery stores. Petrol stations and other retail establishments soon enabled bank card 'debit' payments, relying on automated payments via Automated Clearing house (ACH) technologies through pin number accessible ATM card technology for payment. Eventually convenient and reliable ATM and debit/ACH services won out over worry, in part due to explicit legislation to protect users.

RFID cards (which can be 'read' by nearby hackers), mobile phone apps and Apple Pay transactions may be going through a similar transition as they gain wider acceptance. For example, the Starbucks app enables mobile phone payment at checkout. Unfortunately, its 'automated reloads' feature is not immune to thieves [6].

Automated payments start with general acceptance of card payments, but introduce heterogeneity through differences in different payment processes, and a raft of substantial risks for which people have little information and thus little capacity to manage.

C. Automated Vehicles

Automobiles are a region of adjustment for humans ranging from what in a car can be maintained by an owner to, soon, the driving of the car itself. The computer innards of automobiles and mechanical complexity have greatly restricted the options available for most people to service a car themselves [7]. Soon, driving will become an option as more automated vehicles join traffic, with the Google car operating on Mountain View public roads as of the summer of 2015 [8]. Furthermore, the transition to mixed-autonomy traffic flow will rapidly disrupt established criteria for responsibility and liability, with a corresponding decline in trust of responses by industry and government. Individuals cannot assume responsibility in the long term as passengers in self-driving vehicles, and corporations are also likely to resist absorbing responsibility. More collective solutions must emerge, where responsibility and liability are distributed beyond the single vehicle and manufacturer of the vehicle, initially as a kind of individually funded public transportation, with eventual conversion to a commercial or government service [9].

This process is likely to depend, in part at least, on distributed IoT technologies, with associated issues of acceptance as people gain more control over these heterogeneous technologies. Ethics and trust must adapt to address new capabilities and new situations as we shift to a more automated world.

III. A PERIOD OF MODIFIED TRUST

With mobile payments and semi-autonomous automobiles, people are in a period of adjustment to new technologies, capabilities, and as things change, their trust levels. People incorporate ethics, knowledge and expectations from their cultures, something software generally does not, with conformance to law usually being the closest proxy. As such, a shift from a context that is mostly staffed, run and patronised by people, is increasingly changing to one operated by machines and processes not explicitly programmed to address the ethics and cultural knowledge of the people they serve. The agency (e.g. freedom to make choices from
available options) of the remaining human staff becomes increasingly suppressed as corporate scripts increasingly direct these employees to behave more like the machines [10]. A Starbucks card automatic refill is a 'magical' idea until the algorithm transfers funds to a third party. Employees bound by scripts and processes are not empowered or trained to correct mistakes in other parts of the business. The Starbucks card is governed by an agent/algorithm that operates outside of explicit ethics and cultural knowledge. It may convenient in principle, but humans lose money and time sorting out the obvious mistakes [4] that a design team did not foresee. Starbucks employees, mandated to follow scripts and corporate process rules, may not be empowered or trained to solve digital business problems in a specific physical Starbucks location, unlikely to have the expertise to do what would be required as a solution. Electronic theft is still unfolding in potential [11] and new crimes develop as new technologies provide new capabilities and new means of exploitation. Furthermore, limited jurisdiction of cyber crime enforcement in conjunction with global crime makes catching cyber criminals challenging [12].

Humans are better at determining trust and risk and applying ethics and cultural knowledge than machines—for the moment—and yet when it comes to working on how to solve these 'loophole' problems in development, developers are often unable to do so, as many software companies are either founded by, or populated with, people with deadlines who have limited understanding of cultural knowledge, or even general knowledge, beyond the scope required to produce shipping software [13]. As a result, software (and hardware) development perpetuates automation, often without ethical oversight and/or appropriate cultural knowledge, while attempting to automate processes at an individual level, directly or indirectly requiring humans to bridge the gap.

IV. ALGORITHMIC ADAPTATIONS

A. Humans Bridging the Gap

One solution that is currently being developed to bridge the gap between human needs and algorithmic expressions, takes the form of 'smart' agents. Siri, Assistant, Cortana, and other automated agents attempt to translate requests into information delivered on a mobile phone. Other systems are operated by humans and delivered through texting or other mobile or web services. Software agent mediators are another meet-in-the-middle technology for human interaction with systems, but require trust on the part of the user to enable the system to successfully anticipate and fulfill needs.

B. Magic and Alfred

Apps such as Magic and Alfred are designed to fulfill and pay attention to human needs (with human help). Magic is applied in a much broader context without implying a trusted, getting-to-know you relationship, offering a service that enables a person to text to receive nearly anything desired (as long as it is legal) delivered to them in a single step. People use Magic to book tickets, make reservations, order food, etc. "Magic doesn't actually have a workforce handling the delivery itself. It simply has humans handle the request and figure out what the cost is to fulfill them using regular delivery services. … Magic promises to deliver … as long as they are willing to pay whatever Magic marks the price for delivery to be" [14].

Alfred [15] seems to require more personal, private information than Magic and thus, more trust on the part of the person using it. A write-up of Alfred portrays it as an invisible servant. Alfred is a human agent butler service, composed of an algorithmic and human component, whose actions are routed through a semi-automated app. Alfred could be considered as training for the IoT in that the goal of Alfred is to be an unobtrusive servant, maintaining daily household chores and seamlessly knowing preferences in the home.

The first step of joining Alfred involves filling out a quiz. Alfred (the app) wants to know about the user's preferences. This may include such things type of peanut butter preferred or how they want their supplies arranged [16]. Even the way that Alfred acquires keys is automated:

... you can arrange to give a copy of your keys to your new Alfred, or use the app to scan your key and remotely offer a copy to your Alfred helper. … After that, all that's left is picking a good schedule [16].

The list of what Alfred does is impressive and covers tasks that have to do with household maintenance such as grocery shopping, tailoring, dry cleaning, laundry, prescription pick-up, and special requests. The service pays attention to details such as putting things in their rightful place. The CEO and co-founder of Alfred, Marcella Sapone, summarises this by saying, "We want to take as much off of our members' plates as possible. Other semi-automated on-demand services available today deliver, but stop at the front door" [16]. Alfred provides a trusted environment enabling a human to enter people's homes and complete on-demand 'curb to inside the door' service. As Alfred's mix of human and program collects data and learns what people want and need, as well as their home layout, location, and living preferences, it might eventually become a fully automated AI driven service, with package deliveries dropped in an open window by drone or some other way to complete at least gross motor skill tasks, until automated robots are accurate, available and affordable.

While having an actual robot put things in their place may seem years away, the foundations for the computer algorithms that can enable this activity are beginning now. With a recent focus on indoor home mapping, iRobot, the company that makes Roomba, the small automated robot vacuum cleaner, is planning to "market a robot that can create a map of your home by recognizing and labeling everything in it using a camera and a cloud-based engine". Colin Angle, CEO of iRobot, "describes the maps and their potential use cases as 'the context engine that drives the intent' for future in-home

©2015 Sally A. Applin and Michael D. Fischer unless otherwise noted. All rights reserved.
It is possible to imagine a future where the human patterns from semi-automated services will be converted to personalised algorithms (or AI) that home IoT robots could engage with. Each human in the process for Magic and Alfred, combined with the technology used to create and fulfil the request, is AI helping to gather data for a later date. Right now, we need human intervention to enable the systems and processes to successfully work in a way that enables human agency. With iRobot's intended in-home mapping, this may be closer than we expect.

A. Crowdsourcing Automation

One service crowdsources task fulfilment by offering a hub for people to share automation scripts. Do, "The do it yourself button" app offers a single button that: "empowers you to create your own personalised button with just a tap [and] ... control the world around you with Recipes that connect your button to Philips Hue, Google Drive, Nest ... and hundreds of Channels you use every day" [18]. The "Recipes" that are shared on Do are submitted by other users of the system and available for any other user to add to their own personal cache of "do buttons" to use. Do buttons that have been created include things such as "Neighbourhood watch! If anyone takes an Instagram photo in the area, send an iOS notification" [18] and a way to turn on appliances (if they use the Belkin Wemo switch). Other uses are divided between actions for IoT type devices and services to mute, change or reroute communications or information. The Do service creates a user base that become the programmers and Thing-agents [19] of their own and, by proxy, other's "home", "work" and what the Do service refers to as "essentials" environments.

B. Distribution and the IoT

WASH Laundry was a low-tech business that has been incorporating Microsoft's IoT technology into their coin-operated laundry facilities to become a high-tech laundry business [20][21].

WASH connected its machines to the Internet, then mined data on price and machine use, analysing it with Power BI. ... to fine-tune the price set on each machine in near-real-time, to find the sweet spot between maximising profit and keeping customers from washing elsewhere. With more than a half million machines in use, even the smallest change in price can scale out to big profits ... [22].

WASH is an automated system, that data mines human laundry behavioural habits to maximise profits. The only entity benefitting from this type of autonomous integration is the service provider, which uses analytics to price control laundry for consumers in multiple locales based on their usage data [22]. The implications of this 'just-in-time' pricing for humans when applied to all sorts of services we rely on for our survival are staggering. Uber, the semi-automated ride service offering a 'surge' pricing model of raising rates during events or peak times [23], comes to mind.

VI. AUTOMATION, ETHICS AND ADAPTATION

The autonomous world people want to be building must incorporate ethical principles and balance into its formation. Without mutual trust, it becomes much more challenging to reconcile complex heterogeneous human behaviours in an automated context. As algorithms control more and more of our world, there is a need to understand and regulate their impact, particularly as they create new problems such as the Google cars' inability to avoid rear-end accidents [24].

4. Algorithmic Transparency

One way to achieve at least a start at incorporating ethics, could be through 'algorithmic transparency', a concept described by Cundiff (1984) as a way for researchers to understand simulation and coding complexity applied to continuous systems, "without mathematical training ... [but] have access to general purpose languages ... with which meaningful models can be simulated" [25]. Cundiff proposed using APL (a programming language that uses the multidimensional array in combination with graphic symbols) to express, functional relationships between elements being modelled, to convey the inner workings, establishing a "'glass box' approach in contrast to the 'black box' typically embodied in simulation languages" [25].

Cundiff (1991) extended the concept of 'algorithmic transparency' to "model structure" [26], implementing Interpretive Structural Modeling (ISM)[27] with the "executable notation of APL in direct definition form" [26].

Most relevant for our purposes is Cundiff's approach and insight into the limitations of focusing on functional relationships, which when exclusively applied to human/algorithmic contexts are likely to restrict and limit human agency, options and choices:

System structure is often only indirectly considered in model construction. In focusing on the functional relationships a reductionist orientation may be unconsciously introduced where by important patterns of interaction, intuitively thought to prescribe model behavior, are absent" [26].

Approaches to 'algorithmic transparency' have evolved as we use more data and have thus become more aware of the consequences of our data actions overall as a society. Algorithmic transparency is gaining public awareness as scholarship and publicity on the topic increases. For example, Coddington (2015) examines "Algorithmic Accountability" in the context of journalism and journalistic integrity [28] and Harzog (2015), develops a thorough and useful analysis of robot ethical considerations in "Unfair and Deceptive Robots" [29]. Ashkan Soltani, on being appointed Chief Technologist of the US Federal Trade Commission included a section on algorithmic transparency in his 2015 inaugural blog post, stating, “I hope to expand the
agency’s ability to measure big data’s disparate effects in order to ensure that the algorithms ... afford them the same rights online as ... offline“ [30]. The Organization for Economic Co-operation and Development (OECD) is a fifty-year-old group of thirty-four global nations who come together to "identify problems, discuss and analyse them, and promote policies to solve them” [31]. At their 2014 Global Forum on the Knowledge Economy, OECD discussed "building trust in the data-driven economy" noting that 'algorithmic transparency' raises "complex issues" that were "considered worthy of further examination" [32].

B. Incorporating Ethically-based Frameworks

A framework for referencing embedded ethically-based principles must implement guidelines that are self-consistent, predictable and explicable to establish a consensual basis of use.

Examples include development of religious, legal and professional codes of conduct. In all cases workable implementation of these codes requires a separate process whereby individuals or panels of individuals are empowered and positioned to frequently make judgements or arbitrate. This is not easily scalable to an automated world.

Even if ethically-based 'rules' were converted in some way into algorithms, the result may not be 'ethical algorithms'. Furthermore, there is no framework for ethical behaviour that will be acceptable to all people or organisations in all situations. To embed ethical behaviour into automated processes there must be some means to review outcomes of an algorithm, particularly when interacting with other algorithms embedded in a workplace or a services environment such as the IoT.

Kraus, Sycara, and Evenchik (1998) review a range of mechanisms used by humans for dispute resolution in the context of agent-based systems [33]. Building on these mechanisms, mutually agreed automated external review systems could provide advice (feedback) to an agent as to the likelihood of a specific proposed outcome breaching one or more of several ethically-based frameworks within a given context. Perhaps more effective, review agents could make judgements on real-time outcomes that establish reputations for the services or agents contributing to the outcomes across different ethical domains, maintaining a collective ethical 'scorecard' that users, designers and agents can refer to. This is effectively what happens on the web at the moment when people use a search engine to determine the reputation of an application or library, and could be the basis for yet another example of algorithms 'learning' from humans.

A potentially strong, though indirect, strategy to encourage embedding code to produce more ethical outcomes for algorithms is to focus on improving and expanding the review process rather than the codes or frameworks. Feedback from established review agents, human or otherwise, will create an impetus to develop ongoing correction through normal releases. Review agents would be particularly attractive where there exists an 'ecosystem' of independently sourced algorithms embedded in services and IoT smart environments, developed and used by diverse groups for diverse purposes. Rather than having only fixed embedded routines, some choices could be influenced by submitting proposed outcomes to a review agent that would have access to other proposed outcomes from other active agents in a specific problem context, and would report the likely ethical status of the outcomes within the full context.

This would have the advantage of not blocking many capabilities a priori by users or designers of agent services, while dampening negative ethical impacts for serious cases identified during review, or at least providing more information to the agent or user who may then select more equitable choices.

In all of the examples in §IV the same problem is clear, people are highly heterogeneous, yet with regard to their living experiences have similar but different needs. An automated world will need to be more flexible to be able to integrate with humans. Right now, it is less so. Thus, humans are assisting systems to enable them to become more autonomous and autonomous systems are adapting how they receive input (sometimes requiring humans in the loop) to adapt to humans.

We need to encourage development of ways in which these systems, particularly IoT systems, can enable more designer and user choices, and thus more agency for humans in automated contexts, together with more ethical oversight, yielding improved operation as automated systems become more flexible. El-adaway (2008) discusses using multi-agent simulation as a tool for resolving ongoing disputes in the construction industry [34]. In a similar light, Applin and Fischer propose Thing Theory as a general interface between users and services [19]. Thing Theory builds on the premise of transferring interacting component management to Thing-agents that manage sub-agents within a particular technological context. Using an extensible multi-agent simulation incorporating a specification for each sub-system relating sensors, actuators and associated services, creating a model of how the services might interact with each other, and the contexts that can emerge [19]. By providing contextual simulation services within an automated runtime environment, Thing-agents can identify and offer viable choices to human or autonomous agents, and provide rapid feedback regarding likely outcomes of making different choices. When a sub-agent makes a choice, Thing-agents can facilitate further options for how they can proceed. This will enable more appropriate choices for sub-agents to offer people within automated environments without requiring much assistance from Thing-agents to help along the process. Thing-agents are be a good place to embed ethical review services that can provide individual sub-agents with the contextual information they require to promote ethical outcomes. However, it does require
some yielding of trust within a trusted environment (such as the home) in order for sub-agents to have the data to provide useful options. This also will require ethical consideration and awareness of people's privacy, trust and sense of security.

VII. CONCLUSIONS

We are in a period of adjustment for humans and algorithms. To successfully complete many tasks in the world, there is a growing co-dependence between humans and algorithms. While it may seem that humans are increasingly doing the balance of helping, the existence of algorithms to aid various processes can be very helpful as well. However, what emerges from this new cooperative cycle is a situation whereby ethical adaptations to new technologies and capabilities are available mainly on the human side, together with cultural knowledge and local referenced knowledge. Simultaneously, algorithms are referencing poorly unadapted ethical practices, if at all, in addition to the other data and information these access. The imbalance of ethical responsibility going forward will be a concern, as machines are not yet learning ethical practice from humans, rather they are just learning how humans think of (intelligence) and express (do things as a result of) their thoughts. Going forward, this may create a challenge for humans because as they do 'teach' systems and processes about themselves, there is no receptacle within a machine algorithm at the moment for learning ethics from humans in commercial apps such as Alfred, Magic and WASH.

For most of the past few thousand years, ethical considerations were expressed within the political and religious subsystems that co-evolved with the rest of society. These strongly favoured institutions over individuals. These types of ethical principles can be instantiated as 'rules' that algorithms are able to learn as these can be derived from data and documentation over millennia, but pertain to a world that did not include digital algorithms and nearly ubiquitous computing.

Newer ethical considerations, which take into account new technologies and capabilities, are still being formed [35; 36]. Humans are better at determining ethical boundaries in this adaptive state than algorithms relying on older knowledge. This may be due to the fact that new adaptive ethical knowledge is still being synthesised and refined by humans [35; 36].

Without a means for algorithms to collect new ethical considerations as well as behaviour and intelligence, the reference points for algorithmic ethics will be dated. This is particularly true as society moves from centralised to distributed forms of communication as evidenced by the mobile revolution and complex multiplexed communications networks that result [37].

The risk in full-scale automation is the lack of adaptive ethical knowledge accessible to algorithms based on fixed decision tree pathways in coding that do not account for dynamic human agency choices and expressions. This will leave humans supported by a restricted system. In this mixed-use middle period, we have the opportunity to develop new ways to handle ethics as knowledge that machines can learn.

Future systems will need to be developed to handle the massive amounts of data being generated by multiple, multiplexed devices and the IoT that include allowances for human choices outside the range of fixed algorithmic options, yet are flexible enough to maintain (trusted) privacy and security and to do all of this in an ethical context.

REFERENCES


©2015 Sally A. Applin and Michael D. Fischer unless otherwise noted. All rights reserved.

Citation information: