



Kent Academic Repository

Earp, Brian D., Hannikainen, Ivar, Skorburg, Joshua and Everett, Jim A.C. (2022) *Experimental Philosophical Bioethics of Personal Identity*. In: Tobia, Kevin, ed. *Experimental Philosophy of Identity and the Self. Advances in Experimental Philosophy*. Bloomsbury, London, UK. ISBN 978-1-350-24689-8.

Downloaded from

<https://kar.kent.ac.uk/92038/> The University of Kent's Academic Repository KAR

The version of record is available from

<https://www.bloomsbury.com/uk/experimental-philosophy-of-identity-and-the-self-9781350246898/>

This document version

Author's Accepted Manuscript

DOI for this version

Licence for this version

UNSPECIFIED

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in *Title of Journal*, Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our [Take Down policy](https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies) (available from <https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies>).

Experimental Philosophical Bioethics of Personal Identity

Brian D. Earp, Jonathan Lewis, Joshua August Skorburg,

Ivar R. Hannikainen, and Jim A. C. Everett*

This is the authors' copy of a forthcoming book chapter.

Earp, B. D., Lewis, J., Skorburg, J. A., Hannikainen, I., & Everett, J. A. C. (forthcoming).
Experimental philosophical bioethics of personal identity. In K. Tobia (ed.), *Experimental
Philosophy of Identity and the Self*. London: Bloomsbury.

Abstract: The question of what makes someone the same person through time and change has long been a preoccupation of philosophers. In recent years, the question of what makes ordinary or lay people judge that someone is—or isn't—the same person has caught the interest of experimental psychologists. These latter, empirically oriented researchers have sought to understand the cognitive processes and eliciting factors that shape ordinary people's judgments about personal identity and the self. Still more recently, practitioners within an emerging discipline, experimental philosophical bioethics or "bioxphi"—the focus of this chapter—have adopted a similar aim and employed similar methodologies, but with two distinctive features: (a) a special concern for enhanced ecological validity in the examples and populations studied; and (b) an interest in contributing to substantive normative debates within the wider field of bioethics. Our aim in this chapter is to sample illustrative work on personal identity in bioxphi, explore how it relates to studies in psychology covering similar terrain, and to draw out the implications of this work for matters of bioethical concern. In pursuing these issues, we highlight recent work in bioxphi that includes the perceived validity of advance directives following neurodegeneration, the right of psychologically altered study participants to withdraw from research, how drug addiction may cause one to be regarded by others as "a completely different person", the effect of deep-brain stimulation on perceptions of the self, and the potential influence of moral enhancement interventions on intuitive impressions of a person's character.

Keywords: bioethics, bioxphi, personal identity, advance directives, addiction, moral enhancement

* We are immensely grateful to Joshua Knobe and Kevin Tobia for their invaluable comments and suggestions during the writing process, and we wish to thank Kevin for inviting us to contribute to this collected edition.

Introduction

The question of what makes someone the same person through time and change has long been a preoccupation of philosophers. In recent years, the question of what makes ordinary or lay people (that is, individuals from a wide range of backgrounds, including non-philosophers) judge that someone is—or isn't—the same person has caught the interest of experimental psychologists. These latter, empirically oriented researchers have sought to understand the cognitive processes and eliciting factors that shape ordinary people's judgments about personal identity and the self. Still more recently, practitioners within an emerging discipline, experimental philosophical bioethics or "bioxphi"—the focus of this chapter—have adopted a similar aim and employed similar methodologies, but with two distinctive features: (a) a special concern for enhanced ecological validity in the examples and populations studied; and (b) an interest in contributing to substantive normative debates within the wider field of bioethics (Earp and Demaree-Cotton et al. 2020; Klenk 2020; Lewis 2020a; Earp and Lewis et al. 2021).

Our aim in this chapter is to sample illustrative work on personal identity in bioxphi, explore how it relates to studies in psychology covering similar terrain, and to draw out the implications of this work for matters of bioethical concern. Of course, the boundaries between fields and disciplines are often blurry, and many of the same practitioners are conducting research across these different areas. We are less concerned with whether a given study or line of work counts as an instance of bioxphi than with characterizing bioxphi more generally and exploring some of the ways in which data-driven studies using the methods of cognitive science, moral psychology, and experimental philosophy (x-phi) can inform bioethical argumentation and decision-making.

With this aim in mind, we begin by briefly reviewing classic studies in psychology and bioethics that bear on questions concerning lay judgments about personal identity change versus persistence: that is, the extent to which someone is regarded as the *same person* despite undergoing significant psychological transformation. We will discuss some strengths and limitations of this work with a view to articulating what is new and interesting about bioethics studies investigating similar judgments. Subsequently, we share some candidate strategies for how empirical findings (i.e., descriptive information about how and why people make certain judgments) might appropriately be used in the service of normative arguments: for example, arguments about how personal identity *should* be understood in certain contexts or toward certain ends. We also explore how such empirical information could profitably be employed in the context of practical bioethical decision-making: for example, in helping us to draw conclusions about what *should* be done in various medical situations involving judgments about personal identity change.

In pursuing these issues, we highlight recent work in bioethics that addresses a rich variety of topics touching on personal identity. These topics include the perceived validity of advance directives following neurodegeneration (Earp, Latham, and Tobia 2020), the right of psychologically altered study participants to withdraw from research (Tobia 2016; Dranseika et al. 2021), how drug addiction may cause one to be regarded by others as “a completely different person” (Earp and Skorburg et al. 2019), the effect of deep-brain stimulation on perceptions of the self (Skorburg and Sinnott-Armstrong 2020), and the potential influence of moral enhancement interventions on intuitive impressions of a person’s character (Fabiano 2021). We conclude with some general observations and suggestions for future research.

I. A Broad Overview of Conceptions of Personal Identity Change in Philosophy, X-Phi and Moral Psychology

What determines your personal identity through time and change? Are “you” at age 5 the same person as “you” in old age, at the end of life? If you suffer a serious brain injury that causes you to lose most of your memories—or which permanently alters key aspects of your character—are you still yourself, and if so, in what sense? Suppose that two people could swap bodies (Williams 1970); how could we decide or track who was who? A long tradition in philosophy has debated puzzles like these. A prominent view, typically associated with John Locke, is premised on a diachronic conception of the self as a reasoning and reflective entity that persists over time (Locke 1689/1694 [1985], 2.27.9). For Locke, as he has classically been interpreted, “consciousness” of past actions—via the faculty of memory—is at the heart of personal identity (Reid 1785 [1969]). Broadly, according to this view, Person A who did x at Time 1 can be identified as Person B at Time 2 if B has the same “consciousness” as A, where this implies having a memory of doing x .

Whether this is the correct, or most justifiable, view of personal identity is an open question. As some philosophers have argued, in the context of scientific inquiry, a theory of a given phenomenon is usually considered to be adequate if, among other things, it can explain the relevant data: for example, if it can explain why the contents of a test tube turn green under certain conditions, but red under others (Thompson 1986, 257; Kagan 2001, 47). When one seeks to explain personal identity, however, it is not obvious what the relevant data are that need to be explained. Certainly, many of us have a strong sense that we exist in the form of a person or self that extends through time; and it seems that we live in a world full of other selves that are similarly diachronic. Moreover, experiences of continuity or discontinuity in oneself or in the

selves of others—in the context of close interpersonal relationships, for instance—often matter a great deal. When we “lose ourselves” in periods of mental illness; when a loved one starts to seem like a stranger after developing an addiction (Tobia 2017); when a parent or grandparent with dementia doesn’t recognize us, we may feel disoriented, disenchanted, disturbed. Perhaps *these* are the data that need to be explained: our everyday beliefs, perceptions, and experiences of ourselves and the selves of others through time and change.

Philosophers, including Locke, have developed theories along these lines. However, in doing so, they most often have simply consulted their own intuitions about particular cases—whether real or imagined—that seem to implicate, or raise questions about, the nature of personal identity (e.g., Williams’ “body swap” scenario mentioned above). Commonly, they will ask what is entailed by a given theoretical claim if applied to such a scenario, and then check to see whether the implication seems intuitively right or reasonable to them. If it doesn’t seem reasonable, then they may go back to adjust the theory. But whether this is an appropriate strategy for building a general account of personal identity is unclear.¹ As Machery notes, philosophers who rely on such an approach “are neither inquiring about the actual world, collecting observational data or running experiments, nor examining our best scientific theories to [decide] what determines personal identity” (Machery 2017, 189). Moreover, even if it is granted that philosophers’ intuitions should play some role in theory development, it is doubtful whether these intuitions should be accorded substantially more weight than those of non-philosophers when it comes to providing answers to certain philosophical questions (Machery 2017, 149–184). At the very least, it seems pertinent to know whether the philosophers’

¹ For a critique of the philosopher’s “thought-experiment” approach to questions concerned with personal identity, see Wilkes (1988).

intuitions or associated judgments about cases are idiosyncratic or widely shared among relevant stakeholders.

As Kagan (2001) notes, philosophers have for the most part simply assumed that others—at least, other philosophers—will share their intuitions and judgments about cases. Until relatively recently, this assumption had not been tested much empirically (Machery 2017). However, this started to change in the early 2000s when “experimental philosophers” (Knobe, 2007; Knobe et al. 2012) started to move away from proverbial armchair theorizing, towards investigating how ordinary people think about questions such as what it means to be the same person over time. For example, Blok, Newman, and Rips (2005) asked study participants to imagine that a team of doctors removed the brain of a patient (“Jim”) and destroyed his body. In one scenario, the doctors successfully transplanted Jim’s brain into a new body, so that Jim’s memories were fully preserved: physically, he had a completely different body, but he could still remember everything from before the surgery. In another version of the scenario, all the other details were kept the same except that Jim’s memories were not preserved during the transplantation.

Philosophical debates in the Lockean tradition show that there are sophisticated arguments both for and against someone’s being the same person in a case like this. However, one argument concludes that the brain recipient is the same person as Jim when the latter’s memories are preserved, despite the bodily changes. And that is what the researchers found: participants agreed that the brain recipient was the same person as Jim when his memories remained intact but disagreed when his memories did not survive the operation (Blok, Newman, and Rips 2005; for a replication of the results of Blok, Newman and Rips’ third-person investigation, see Nichols and Bruno 2010). It seems, then, that ordinary people—not only those

with advanced philosophical training—intuitively regard autobiographical memories as being at least one important contributing factor to personal identity.

Other factors matter as well. In a recent, influential line of work, it has been demonstrated that *moral attributes*, even more so than memories, are widely judged to be at the core of personal identity (Strohming and Nichols 2014; Prinz and Nichols 2016). In these studies, researchers gave participants a list of traits and asked them to imagine to what degree a change to each trait would influence whether someone was still the same person. In what is now known as the *moral self effect*, changes to a person’s moral qualities were consistently judged to be the most identity-disruptive factor. As Strohming and Nichols concluded, “moral traits are considered more important to personal identity than any other part of the mind” (Strohming and Nichols 2014, 168). Furthermore, the *direction* of moral change also matters: evidence suggests that improvements to moral character, compared to deteriorations, tend to be seen as less threatening to personal identity persistence (Tobia 2015).

So far, these studies have overwhelmingly involved “WEIRD” participants (Henrich, Heine & Norenzayan 2010), so that the range of cultural and other demographic contexts across which similar results might obtain is not yet known (but see, e.g., Dranseika et al. 2021).² Nevertheless, in the meantime, philosophers have begun grappling with the (meta)philosophical implications of experimentally derived findings regarding people’s judgments of personal identity (e.g., O’Neill and Machery 2014; Fisher 2015; Knobe 2016; Machery 2017; Nado 2021). In line with this approach, but focusing specifically on philosophical bioethics, we will now explore some of the implications of such findings for the burgeoning discipline of bioxphi.

² For a broader debate regarding whether intuitions vary across demographic groups in general, see Knobe (2019a); Machery and Stich (2019); Knobe (2019b).

II. From X-Phi and Moral Psychology to Bioxphi

The studies surveyed above, and others like them, have done much to increase our understanding of how ordinary people think about personal identity change or persistence, albeit sometimes in relation to unusual cases. Typically, these studies rely on vignette-based designs that hold everything constant between experimental conditions apart from a specific feature—such as a given trait or personal attribute—that is expected to make a difference to participant judgments. Although questions have been raised as to whether this contrastive-vignette technique (CVT) should be used so exclusively (Mihailov, Earp, and Hannikainen 2021; Earp and Lewis et al. 2021), an advantage of the approach is it allows researchers to zero in on well-defined aspects of the self that might change as a result of some imagined intervention, so that fine-grained discriminations can be made between potential factors shaping participant judgments (Reiner 2019).

When applied to questions of personal identity, such CVT studies oftentimes have employed far-fetched, even science-fiction-like examples: brain transplants, body-switches, magic pills, time machines, reincarnation, and the like. This likely has to do with a typical aim of such studies, which is to validate or refute the premises of arguments made from the armchair. As Machery (2017, 113–116) argues, such “armchair” philosophizing often relies on cases that are deliberately “unusual” in order to achieve certain theoretical goals. However, researchers increasingly are showing an interest in how cases might be adjusted to be more realistic (or to track features of the world that are more representative of everyday situations).

As alluded to previously, this includes researchers in bioxphi, an emerging discipline which uses the tools of x-phi, moral psychology, and cognitive science to investigate topics in bioethics with a heightened emphasis on ecological validity. In a bioethical context, judgments about whether, to what extent, or in what sense someone is the same person despite having undergone various changes often matter for real-life decisions about how someone should be treated: for example, in the context of high-stakes decisions about what healthcare a person should receive as a part of their end-of-life care. Understanding the factors that influence these personal identity-related judgments in more realistic, ethically charged situations may thus be both theoretically and practically relevant (but see Shoemaker 2010).

Some work in the more established tradition of moral psychology has already taken a turn towards greater realism. For example, Strohminger and Nichols (2015) studied judgments of identity persistence made by actual family members of patients with different neurodegenerative diseases. The authors sought to determine the extent to which patients would be seen as a different person as a function of three types of changes: changes to moral faculties (as in some cases of frontotemporal dementia), changes to memories (as in Alzheimer's disease), and changes to physical motor functions (as in amyotrophic lateral sclerosis). Mirroring results from earlier abstract thought experiments, Strohminger and Nichols found that family members of patients with frontotemporal dementia—the condition most strongly associated with moral change—saw the patient as more of a different person compared with family members of patients with other forms of neurodegeneration (in line with the moral self effect). Strohminger and Nichols also found that moral changes led to more severe deterioration in the relationships between patients and family members. Reflecting on these results, the authors derived normative recommendations:

While loss of identity may be feared as an undesirable clinical outcome unto itself, the present research highlights that identity deterioration has significant downstream consequences for healthy relationships. ... *Future therapies ought to be aimed at—and take into account—preserving moral function*, a previously unappreciated factor in the well-being of patients and their families. (Strohminger and Nichols 2015, 1477, emphasis added)

The bioethical analysis here is straightforward. First, it is noted that certain kinds of neurodegeneration seem to impair both personal and interpersonal well-being through their impact on moral functioning. Since one of the main normative commitments of healthcare is to mitigate threats to well-being that stem from disease or disability—a premise Strohminger and Nichols implicitly accept—then, assuming sufficient resources, among other relevant considerations (e.g., ranking of research and funding priorities based on likely success, effectiveness, distributive justice, and so on), it follows that therapies *should* be developed to address this mediating factor.

The moral self effect on judgments of personal identity also applies to the case of addiction (Earp and Skorburg et al. 2019), another real-world issue that can have major implications for personal and interpersonal well-being. Over a series of experiments involving contrastive vignettes inspired by real-life stories—and designed to be as believable as possible—three of us explored whether participants would judge that someone who became addicted to drugs would thereby become, in some sense, a different person; and, if so, whether this judgment would be driven by perceived changes in the individual’s moral qualities (Earp and Skorburg et

al. 2019). In an initial experiment, U.S. participants judged a character who became addicted to drugs as being closer to “a completely different person” than “completely the same person,” while subsequent experiments revealed that these judgments were indeed driven by a perceived worsening of the moral character of the drug user. In particular, the user was seen as having drifted away from their (presumably morally good) “true self” (ibid.).

Could findings like these have normative implications? Although we could only speculate based on our initial results, we suggested that public health messages framed around (re)discovering one’s true self might be especially desirable in the case of addiction (for related work, see Schlegel and Hicks 2011; Schlegel et al. 2009; 2011). Another possibility is that campaigns aimed at *dissociating* drug use or addiction from negative moral character judgments could be beneficial (Hart 2021), for example, by inclining the public to believe that individuals who are dealing with addiction are still the same people after all. In either case, the underlying bioethical reasoning would be similar to that employed by Strohming and Nichols (2015). In short, if drug addiction harms individuals and relationships, and if some of this harm comes from the perception that drug addiction undermines the identity of the user, then, *ceteris paribus*, measures to address that perception should be pursued.

In the following section, we will look at other ways in which empirical results—derived from bioxphi studies—can help inform normative conclusions about bioethical issues involving judgments about personal identity. First, we will describe some general strategies for reaching normative conclusions from premises that include empirical information about the mind. Second, we will discuss these strategies in relation to specific examples drawn from the recent bioxphi literature.

III. From Empirical Studies to Normative Conclusions: Four Recent Strategies from the Bioxphi Literature

We have argued that bioxphi studies should be designed with at least two aims in mind. The first aim is scientific: it is to understand the cognitive processes and eliciting factors that shape morally relevant judgments in the real world, so that we might build theoretically justified and descriptively accurate models of the—realistically situated—moral mind (Mihailov, Earp, and Hannikainen 2021) The second aim is normative: it is to harness these models and associated findings to help reach ethically warranted conclusions in bioethics (e.g., regarding public health policy or clinical decision making) (Earp and Lewis et al. 2021).

In a recent paper (Earp and Lewis et al. 2021), some of us drew on examples from the burgeoning bioxphi literature to outline four main strategies for reaching normative conclusions from premises that include empirical content about the mind: *parsimony*, *debunking*, *triangulation*, and *pluralism*. Here is a brief description of each strategy, followed by an illustrative example of how it might be applied to an ethical question involving judgments about personal identity:

1. **Parsimony.** If relevant stakeholders consistently make a judgment p which encodes moral claim M , then M has *prima facie* normative weight (inspired, in part, by consultative approaches to empirical bioethics as detailed in Davies, Ives, and Dunn 2015).
2. **Debunking.** A strategy seeking to show that a judgment or moral claim is unreliable using the following argument structure (inspired by Mukerji 2019):

(P1) Judgment *p* is the output of a psychological process that possesses the empirical property of being substantially influenced by factor F.

(Empirical premise)

(P2) If a judgment is the output of a psychological process that possesses the empirical property of being substantially influenced by factor F, then it is *pro tanto* unreliable. (Bridging normative premise)

(C) Judgment *p* is *pro tanto* unreliable.

3. **Triangulation.** Divergence among the judgments of various groups of experts and/or between expert and lay judgments requires the following: adjusting, pruning, or supplementing the normative conclusions derived from either expert or lay judgments in order to accommodate: (1) the normative implications of the opposing views; and (2) normative considerations derived from, for example, ethical or legal principles, background theories, morally relevant facts and/or the best arguments for a normative position in the relevant expert literature (inspired, in part, by the model of reflective equilibrium, and, in part, by the concept of triangulation in philosophy of science as detailed in Kuorikoski and Marchionni 2016).
4. **Pluralism.** In cases where expert and lay stakeholders hold conflicting, yet *pro tanto* reliable, judgments, or where multiple and independent communities each reveal persistent disagreement between two or more conflicting yet *pro tanto* reliable judgments, these judgments could all have comparable normative weight; associated normative conclusions may justifiably be agent- or community-relative

and/or preference-sensitive (inspired by the model of Shared Decision Making in clinical contexts as discussed in, e.g., Elwyn, Tilburt, and Montor 2013; Lewis 2020b).

A. Illustration #1: Parsimony

In a recent study (Earp and Hannikainen et al., forthcoming), some of us asked participants to consider the case of a late-stage dementia patient who currently lacks autonomous decision-making capacity. We told participants that the patient had previously, autonomously signed an advance directive (AD) instructing the withholding of treatment just in case she fell ill under certain conditions. When those conditions are fulfilled, however, the patient seems to be living a happy life apart from the illness, which is described as easily curable. We asked research participants whether the AD should be followed, and how much they agreed that the dementia patient was “still her true self” despite the effects of the disease on her personal characteristics.

We found that, among the participants who judged that the patient was still her true self, almost all of them felt that the AD should be ignored or overridden that is, they said that the patient should be treated under the very conditions identified in the AD as sufficient grounds to withhold treatment.

Some bioethicists argue that the preferences of an autonomous individual at Time 1—as recorded in an AD—should determine how a non-autonomous individual at Time 2 is treated, just so long as the individual at Time 1 is the same person as the individual at Time 2 (Buchanan 1988; Buchanan and Brock 1990; see also Dworkin 1993; for a critical discussion of the debate, see Furberg 2012). Given such a bioethical argument, when we consider the relationship between intuitions or judgments regarding an AD and judgments of personal identity change, the

empirical result we obtained might seem surprising. In effect, participants indicated that the patient's AD should be ignored or overridden despite her having undergone substantial personal change through neurodegeneration—precisely the situation an AD is meant to cover.

However, we can, perhaps, make sense of the position adopted by these participants as follows: if an individual is “still her true self” at Time 2 (notwithstanding the loss of autonomy), then her Time 1 preferences as recorded in an AD should *not* necessarily determine how she is treated at Time 2.

Of course, much more work will be needed to clarify this finding,³ and to determine how robust it is across different measures or operationalizations. It may turn out that the apparent tension between the judgments of ordinary research participants about this case, and those of at least some philosophers and bioethicists, is illusory, and that—if the case were described differently, for example—there would be perfect agreement about what should be done in terms of treatment. But let us assume for the sake of argument that the finding does hold up in future studies. According to the parsimony strategy for reaching normative conclusions from empirical findings about people's morally relevant judgments, we should, in this case—as a *prima facie* matter of ethical analysis—conclude the following: that dementia patients who still seem like their “true selves” *should in fact be treated* under the stated conditions, even if this conflicts with the person's previously expressed preferences as recorded in an advance directive.

An immediate objection to this approach is that it may seem to reduce bioethical reasoning to a popularity contest. The mere fact that a majority of research participants—even assuming that their views are representative of some wider population of relevant stakeholders—reach a given moral conclusion about an identity-related case does not entail that this is the

³ See Earp and Hannikainen et al. (forthcoming) for details of a follow-up study.

correct, most reasonable, or (otherwise) most justifiable conclusion. We would not argue otherwise. The claim implied by the parsimony strategy is not that we should rely directly on *argumenta ad populum* to reach normative conclusions in bioethics. Rather, the claim is that the experimentally robust, consistent judgments of ordinary people (or other relevant stakeholders) about a given case or set of cases is one factor that counts in favor of the moral statement or conclusion embedded in, or entailed by, those very judgments.

However, this factor alone will never be normatively decisive. For example, if people's judgments can be shown to be unreliable using a debunking strategy (see below), or if their moral concerns are plausibly outweighed by other, competing moral considerations (triangulation), then it could be reasonable to discount those judgments when deciding what to do. Simply put, the parsimony approach puts the burden of proof on those who would argue that *no* normative weight should be assigned to the consistent judgments of relevant stakeholders about a given moral issue (Earp and Lewis et al. 2021).

B. Illustration #2: Debunking

The parsimony strategy is parsimonious in that it provides the simplest possible model for deriving normative content from descriptive information about people's moral judgments. It holds that these judgments should, defeasibly, be given *some* normative weight. However, suppose a researcher is skeptical about the normative conclusion consistently reached by a group of stakeholders about some case. The researcher might want to engage in a debunking strategy as outlined above, potentially using experimental methods to test whether the moral judgments in question are the output of a normatively unreliable factor or process. For example, the researcher might want to see if the judgments are susceptible to framing effects that should have no moral relevance to the issue at hand (but see Demaree-Cotton 2016); or whether, perhaps, they are

influenced by factors that are themselves morally objectionable, such as racial bias or sexist attitudes.

Consider the moral self effect by way of illustration. As discussed previously, this effect refers to the tendency of participants to judge that someone who undergoes a significant moral change—especially when this change involves moral decline—becomes in some sense a different person. As Tobia (2016) argues, whether someone is seen as undergoing a moral improvement versus a deterioration will, therefore, often have different effects on judgments about their personal identity persistence; and these effects, in turn, may influence judgments about matters of bioethical concern. As an example, Tobia refers to an ongoing debate in bioethics concerning the moral acceptability of cognitive enhancement.

Suppose that, in a society shaped by sexist values, a group of participants consistently judged that a certain cognitive enhancement procedure (geared toward boosting one's intelligence and agency) was less ethically worrisome if pursued by men than if pursued by women; and suppose this judgment was rooted in the perception that the procedure would be more threatening to the personal identity of women. Suppose we grant that if a procedure is more threatening to the personal identity of members of one group compared to another, then this is an ethically valid reason to object more strongly to its use in the relevant group.

Given the moral self effect, we can imagine a possible explanation for this hypothetical discrepancy in judgments regarding the impact of the agency-boosting procedure on the personal identity of women versus men: the procedure might be thought to cause greater *moral deterioration* in women. This, in turn, could be due to a sexist societal conception of the “moral woman” as passive, submissive, or (otherwise) lacking in agency (Garcia 2021). Our skeptical researcher, then, might undertake an experimental study to see whether manipulation of this

sexist attitude has the predicted downstream effect on participants' judgments (i.e., about the perceived impact of the procedure on personal identity), and, hence, on the relative acceptability of its use by women as opposed to men.

Of course, the very same stakeholders who might judge that it is worse for women than for men to engage in cognitive enhancement might disagree with the researcher's normative premise, namely, that it is sexist (and therefore wrong) to believe that women ought to be more "submissive" than men. This raises a simple but important point: the debunking strategy for drawing normative implications from empirical findings about people's judgments always involves one or more normative premises which may themselves be a matter of contestation. As a result, the form of the argument is necessarily conditional: *if* you agree with this normative premise (e.g., that moral judgments rooted in sexist attitudes ought to be discounted, and that the attitude in question really is sexist), and *if* the empirical data suggest that this particular moral judgement is rooted in said attitude, then you should discount the judgment accordingly (i.e., assign less normative weight to it).

The point about potential normative disagreements between various stakeholders, and/or disagreements in moral judgment regarding particular cases, leads to the triangulation strategy, described next.

C. Illustration #3: Triangulation

The triangulation strategy holds that divergence among the judgments of various individuals or groups of stakeholders requires the following: adjusting, pruning, or supplementing the normative conclusions derived from the competing positions in order to accommodate: (1) the normative implications of the opposing view(s); and (2) normative considerations derived from, for example, ethical or legal principles, background theories, morally relevant facts, and/or the

best arguments for a normative position in the relevant expert literature (Earp and Lewis et al. 2021). To illustrate this strategy, we will refer to recent work conducted by Dranseika and colleagues (2021) on judgments about the right of study participants to withdraw from research after undergoing significant personal change.

The study builds on earlier findings from Tobia (2016), who asked participants to consider the case of a man who enrolls in a research study and then suffers a terrible accident, as a result of which he experiences (depending on the experimental condition) either moral improvement or deterioration. Tobia asked participants whether the morally changed man should be allowed to have the research study data, which had already been collected before the moral change, destroyed. Participants tended to judge that the morally deteriorated research subject should be denied the right to destroy his data, whereas the morally improved research subject retained the right.

In addition to probing the intuitions of ordinary citizens about Tobia's case, Dranseika and colleagues' cross-cultural replication study included a group of lawyers, whose judgment is especially consequential in real-world decisions involving legal rights. Potentially owing to a legal concept of personhood, lawyers from different countries revealed a distinct pattern of moral judgments: though they were still susceptible to the basic effect (i.e., asymmetry in judgments between conditions), lawyers tended to ascribe identity persistence overall, regardless of the subject's direction of moral change. Thus, at a broad level, Dranseika and colleagues find that (one of)⁴ the lay intuition(s) regarding personal identity is morally laden in a wide range of

⁴ See Knobe (in press) for a discussion of personal identity as a "dual character" concept, suggesting that, even among ordinary people, there may be different concepts of personal identity at play.

cultures, whereas among lawyers, the concept of identity is not equally affected by the same moral considerations.

Put another way, laypeople demonstrated a paternalistic attitude toward the hypothetical research participants, willing to deny that individuals who underwent moral deterioration retained the right to withdraw from research. This attitude conflicted with the legally informed opinion of lawyers, who were more likely to protect the right to withdraw from research even for subjects who had undergone abrupt moral deterioration. When adopting the triangulation strategy, we would begin by acknowledging the divergence in judgments between stakeholder groups (in this case, lawyers and ordinary citizens); and, at least tentatively, assume that the application of either judgment is not straightforwardly amenable to debunking (i.e., both judgments are *pro tanto* reliable). The triangulation approach presents a possible pathway toward prescriptive insights in circumstances like this.⁵

In relation to the above example, it may be the case that the judgments of lay participants toward personal identity change drive their judgments regarding the rights to withdraw from research. If there is good reason to think that a legal conception of personal identity or associated rights should control in some context, the public might need to be educated so as to clarify, or remedy confusions in, their conception as it bears on the legal right.

⁵ However, as with the debunking strategy, the path to a single normative conclusion about what to do in such an instance is not straightforward. When attitudes are, for example, diametrically opposed, triangulation may not be as simple as adjusting or supplementing two or more *pro tanto* reliable judgments in order to achieve some sort of compromise. Instead, one might need to employ a model of “wide” reflective equilibrium (for discussion, see Earp and Lewis et al. 2021). The aim here would be to achieve as much coherence as possible (given the data and information available at the time) between not only the conflicting judgments in question, but also, for example, background theories, legal or philosophical principles, morally relevant facts, etc. (Cath 2016).

Alternatively, consideration of lay judgments might lead us to conclude that the legal concept of identity is not fit for purpose when applied to the question of participants' right to withdraw in research contexts. In such a case, we might advocate stakeholder-motivated "conceptual engineering" in law, as has been done with respect to certain concepts in clinical medicine (see, for example, McMillan 2018; Lewis 2020a).

D. Illustration #4: Pluralism

The pluralist solution to normative inference differs fundamentally from the previous three approaches. Let's suppose that existing empirical studies reveal substantial disagreement on questions of personal identity, and, further, that follow-up evidence does not provide a reason to deem one or the other intuition or judgment unreliable or normatively suspect (i.e., in a way that would help to establish a debunking argument). Furthermore, perhaps additional normative considerations (justice, autonomy, etc.) that philosophers might use to arbitrate between competing judgments fail, in this case, to support one normative view over the other. In circumstances like this, empirical evidence may lend itself to a different normative inference strategy: the pluralist approach. Again, this approach holds that in certain cases of two or more conflicting, yet *pro tanto* reliable, judgments between stakeholders, multiple judgments may be reasonably assigned comparable normative weight.

Here is an example. An important issue involving judgments about personal identity persistence is the set of criteria that should be used to determine whether someone has *died*. Consider a recent study by Neiders and Dranseika (2020), in which they asked participants to express their preference for the stage in the process of dying at which their death should be declared by healthcare practitioners. To support conditions for the *prima facie* normative credibility of participants' judgments, three of the stages from which participants could choose

mimicked three expert conceptions of death in the bioethics literature, specifically, higher-brain death, whole-brain death, and cardiopulmonary death.

In terms of expert judgment, there is more than one medically and ethically reasonable option, which suggests that decisions regarding death determination are, at least in part, preference- or goal-sensitive (Veatch and Ross 2016). In other words, which concept of death we should use may depend on our values or aims. Furthermore, the data gathered by Neiders and Dranseika (2020) reveal widely differing, but stable preferences concerning death determination criteria among study participants. Taken together, the expert view and the study's findings support a pluralistic approach to death determination, according to which individuals should be allowed to choose—within reason—what criterion will be applied in the case of their own death.

IV. Two More Examples

In the previous section, we reviewed four strategies for reaching normative conclusions in bioethics (in part) from experimental findings about stakeholders' moral judgments concerning identity-related cases. However, the examples we used cover just some of the topics being studied in this area. To give a sense of the wider set of issues of potential interest to bioxphi, we will explore two more interventions that have received considerable bioethical attention, and which have potential implications for personal identity: deep brain stimulation and moral enhancement.

A. Deep brain stimulation

In recent work, Skorburg and Sinnott-Armstrong (2020) explored a bioxphi approach to personal identity and Deep Brain Stimulation (DBS). DBS involves a surgically implanted, battery-operated device which delivers targeted electrical stimulation to a specific brain region. The most

common application of DBS is in the treatment of Parkinson's disease. At least 200,000 people worldwide live with neural implants for Parkinson's and related conditions.

Many patients have reported significant changes to their identity after DBS. For example, Schüpbach and colleagues (2006) found that in a population of 29 patients undergoing DBS for Parkinson's, most reported improvements in motor symptoms, activities of daily living, and quality of life. However, marital conflicts were reported in 17 of the 24 married couples and over half were unable to return to professional activities after surgery. Two-thirds expressed feelings of estrangement and unfamiliarity with themselves, saying things like "I don't feel like myself anymore," and "I haven't found myself again after the operation" (ibid., 1813). In another study, patients who underwent DBS for Obsessive Compulsive Disorder or Treatment Resistant Depression reported feeling like themselves but without depression, and feeling "back to" themselves, "back to sort of a baseline" (Klein et al. 2016, 144).

These kinds of case reports and many others like them have raised the question of whether, or in what sense, DBS changes a person's identity. We do not take a stand on this question here (but see Gilbert, Viaña, and Ineichen 2018, for some important criticisms of the claim that DBS does threaten identity). Instead, we want to point out that perhaps the central issue in these ongoing debates is, according to Witt and colleagues (2013, 501), "an explication of what we mean when judging that someone has become 'another person.'" This is, of course, a central aim of the various philosophical and experimental research strategies described in preceding sections.

There are (at least) two upshots here where future work in bioxphi is specifically concerned. First, these real-life examples of post-DBS changes provide ideal kinds of cases to improve the ecological validity of the oftentimes far-fetched thought experiments which, as we

have already observed, tend to dominate theoretical and x-phi approaches to personal identity (e.g., brain transplants, body switches, magic pills, reincarnation, etc.). Second, the bioxphi of personal identity could guide the construction of new forms of assessment of technologies like DBS. Given the moral self effect, for example, assessment of the moral attributes of patients before and after undergoing neurostimulation could shed light on what leads patients or their families to say, “they’re not the same person anymore.”

B. Moral enhancement

The bioxphi of personal identity is also relevant to ongoing debates about moral enhancement. In short, one argument holds that human moral psychology is ill-equipped to deal with the breakneck speed of recent technological developments and the pressing existential risks that they engender (climate change, nuclear weapons, artificial intelligence, and so on). As a result, proponents argue that (biomedical) enhancement of moral traits may be required in order to adequately address these and other issues (Persson and Savulescu 2021). Setting aside the many nuances in these debates, it is worth highlighting that worries about identity change have figured prominently in the enhancement debates from the outset (e.g., Douglas 2008). More specifically, the worry is that enhancing some trait(s) might threaten the psychological continuity often thought to be constitutive of personal identity.

As we noted above, much recent work in x-phi has treated these kinds of questions as empirical ones. The evidence we described on the moral self effect sheds light on which traits are more likely to be viewed as constitutive of identity, how the direction of change in these traits might alter judgments of identity change, and how different contexts shape these judgments. A recent paper by Fabiano (2021) explicitly connects this research program with the moral enhancement literature. He argues that one way to preserve, or even strengthen, the kinds of

psychological relations constitutive of personal identity is “to focus on enhancing virtue because virtue and personal identity are often deeply connected. The relationship between virtue and personal identity is evidenced by the fact that many concepts of virtue are intimately related to personal identity” (ibid., 95). Fabiano claims that much of the evidence we discussed above supports this conclusion. Whether or not virtues are indeed an appropriate target of moral enhancement, we contend that bioxphi approaches to personal identity may meaningfully contribute to these longstanding debates in bioethics.

Conclusion and Future Directions

In healthcare or biomedical research contexts, judgments about whether a patient or participant is the same person matter, and they have real-world implications that rarely arise in armchair contexts. Rights, liberty, consent, autonomy, specific care decisions, social support, and educational interventions are all potentially at stake. We do not claim that bioxphi can establish or even seeks to establish a final theory of personal identity, including how personal identity should be understood in various contexts or with respect to certain ends. Nor do we claim that bioxphi can provide definitive answers to how patients or research participants *should* be treated when changes to their personal identity are judged to have occurred.

Our vision for the bioxphi of personal identity is more modest. Through examples discussed both here and in the burgeoning literature, we can see that this emerging discipline, at least relative to traditional approaches to analytic philosophy and x-phi, is in a better position to support *ethical* decision making in healthcare-related contexts. This is because bioxphi seeks to

understand the factors that influence judgments regarding personal identity in real-world settings: factors that include addiction, the effects of neurosurgical procedures such as DBS, and the decline of cognitive function associated with certain disorders. And it seeks to understand these factors in such a way that the resulting empirical data can be used to support normative arguments.

In attempting to establish, for example, whether stakeholder judgments are *pro tanto* (un)reliable, bioxphi approaches to personal identity can bring empirical data into the service of reaching normative conclusions that are of significance to healthcare practice and policy. As we have seen, these conclusions can relate to, for example, the conditions under which advance directives should be respected or ignored, the specification and enactment of the right to withdraw in cases involving psychologically altered research participants, and the development of therapeutic interventions for people with substance abuse disorders.

The strategies that we have discussed for developing normative inferences from premises that include empirical content are not exhaustive nor should this paper be viewed as defending any single approach. Furthermore, whilst bioxphi studies have relied on the employment of the CVT, proponents of x-phi have argued that we should seek to employ the full range of experimental methods used in the psychosocial sciences, in combination with non-experimental approaches, such as interviews, qualitative studies, studies of linguistic corpus data, and anthropological work (O'Neill and Machery 2014; Alfano, Loeb, and Plakias 2018; Nado 2021). We think this lesson can be applied to bioxphi as well.

Where questions of personal identity change are concerned, the focus of research so far has been on psychological continuity, memories, the self-contained body, or the metaphysically distinct person. In broad terms, the focus has been on “brains and bodies” (Tobia 2016). But, as

more theoretically oriented (yet still empirically informed) practitioners of moral psychology have observed, both the self and identity are not only diachronic, but also relational (e.g., Mackenzie and Stoljar 2000; Christman 2004; Meyers 2005; Earp and McLoughlin et al., 2021). Therefore, rather than focusing so exclusively on a liberal conception of the self as a rational, reflective, and unitary entity, future bioxphi research on personal identity could consider more *relational* approaches to identity that take into account how the self is shaped or even constituted by social and interpersonal relations, structures, and processes.

Toward a relational turn in bioxphi of personal identity?

Individuals are socially embedded. When social norms and interpersonal ways of relating are internalized, “this contributes to the individual’s identity, and thus the identity of the self-as-social is invested in a community and its cultural heritage” (Meyers 2005, 29–30). Ultimately, this suggests that judgments regarding a change in one’s personal identity or that of a close friend or family member can be causally affected by interpersonal relations and the social environment. The psychologically continuous, reflective, and embodied self cannot be isolated from its social-relational situatedness (Lewis 2021).

To that extent, interpersonal relations and social structures can support or impair not only perceptions of personal identity continuity, but also cognitive and moral capacities, character traits, and other qualities relevant to such perceptions (Tobia 2016; Lewis 2021; Veit et al., in press). Doubts that an individual is the same person can arise “not only because of internal factors such as illness, depression, addiction, anxiety and fatigue, but also on the basis of external causal factors such as brainwashing, internalised oppression, stigmatisation, disrespect or inappropriate normative expectations resulting from previous encounters with overly paternalistic, demeaning or pressurising institutional practices” (Lewis 2021, p. 20).

These effects on judgments regarding personal identity are contingent rather than necessary, meaning that whether a specific individual experiences these effects will, ultimately, depend on their psychological states and dispositions, which, in part, constitute their identity. As a concluding thought, we suggest that the bioxphi of personal identity should explore effects of interpersonal relationships on judgments regarding personal identity change and the moral relevance of such relational factors.

References

- Alfano, M., D. Loeb, and A. Plakias (2018). “Experimental Moral Philosophy,” in E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Available online: <https://plato.stanford.edu/archives/win2018/entries/experimental-moral/> (accessed October 26, 2021).
- Blok, S., G. Newman, and L. J. Rips (2005). “Individuals and Their Concepts,” in W. Ahn, R. L. Goldstone, B. C. Love, A. B. Markman and P. Wolff (eds.), *Categorization Inside and Outside the Laboratory: Essays in Honor of Douglas L. Medin*, 127–149, Washington, DC: American Psychological Association.
- Buchanan A. (1988). Advance Directives and the Personal Identity Problem, *Philosophy & Public Affairs*, 17 (4): 277–302.
- Buchanan, A. and D. Brock (1990). *Deciding for Others. The Ethics of Surrogate Decision Making*, Cambridge: Cambridge University Press.
- Cath, Y. (2016). “Reflective Equilibrium,” in H. Cappelen, T. S. Gendler, and J. Hawthorn (eds.), *The Oxford Handbook of Philosophical Methodology*, 213–30, Oxford: Oxford University Press.

- Christman, J. (2004). Relational Autonomy, Liberal Individualism and the Social Constitution of Selves, *Philosophical Studies*, 117 (1/2): 143–164.
- Davies, R., J. Ives, and M. Dunn (2015). A Systematic Review of Empirical Bioethics Methodologies, *BMC Medical Ethics*, 16 (15): 1–13.
- Demaree-Cotton, J. (2016). Do Framing Effects Make Moral Intuitions Unreliable?. *Philosophical Psychology*, 29 (1): 1–22.
- Douglas, T. (2008). Moral Enhancement, *Journal of Applied Philosophy*, 25 (3): 228–245.
- Dranseika, V. (2017). On the Ambiguity of ‘the Same Person,’ *AJOB Neuroscience*, 8 (3): 184–186.
- Dranseika, V., I. R. Hannikainen, P. Bystranowski, B. D. Earp, K. P. Tobia, G. Almeida, G., ... T. Żuradzki (2021). Personal Identity, Direction of Change, and the Right to Withdraw from Research, unpublished manuscript.
- Dworkin, R. (1993). *Life’s Dominion. An Argument about Abortion and Euthanasia*, London: Harper Collins.
- Earp, B. D., J. Demaree-Cotton, M. Dunn, V. Dranseika, J. A. C. Everett, A. Feltz, ... K. P. Tobia (2020). Experimental Philosophical Bioethics, *AJOB Empirical Bioethics*, 11 (1): 30–33.
- Earp, B. D., I. Hannikainen, S. Dale, and S. Latham (in press). “Experimental Philosophical Bioethics, Advance Directives, and the True Self in Dementia,” in A. De Block and K. Hens (eds.), *Experimental Philosophy of Medicine*, London: Bloomsbury.
- Earp, B. D., S. R. Latham, and K. P. Tobia (2020). Personal Transformation and Advance Directives: An Experimental Bioethics Approach, *The American Journal of Bioethics*, 20 (8): 72–75.

- Earp, B. D., J. Lewis, V. Dranseika, and I. Hannikainen (2021). Experimental Philosophical Bioethics and Normative Inference, *Theoretical Medicine and Bioethics* (online ahead of print). Available online: doi: 10.1007/s11017-021-09546-z.
- Earp, B. D., K. L. McLoughlin, J. T. Monrad, M. S. Clark, and M. J. Crockett (2021). How Social Relationships Shape Moral Wrongness Judgments, *Nature Communications*, 12 (5776): 1-13.
- Earp, B. D., J. A. Skorburg, J. A. Everett, and J. Savulescu (2019). Addiction, Identity, Morality, *AJOB Empirical Bioethics*, 10 (2): 136–153.
- Elwyn, G., J. Tilburt, and V. Montori (2013). The Ethical Imperative for Shared Decision-Making, *European Journal for Person Centered Healthcare*, 1: 129–131.
- Fabiano, J. (2021). Virtue Theory for Moral Enhancement, *AJOB Neuroscience*, 12 (2–3): 89–102.
- Fisher, J. C. (2015). Pragmatic Experimental Philosophy, *Philosophical Psychology*, 28 (3): 412–433.
- Furberg, E. (2012). Advance Directives and Personal Identity: What is the Problem?. *The Journal of Medicine and Philosophy*, 37 (1): 60–73.
- Garcia, M. (2021). *We Are Not Born Submissive: How Patriarchy Shapes Women's Lives*, Princeton University Press.
- Gilbert, F., J. Viaña, and C. Ineichen (2018). Deflating the “DBS Causes Personality Changes” Bubble, *Neuroethics*, 1–17.
- Hart, C. L. (2021). *Drug Use for Grown-Ups: Chasing Liberty in the Land of Fear*, New York: Penguin.

- Henrich, J., S. J. Heine, and A. Norenzayan (2010). Most People are not WEIRD, *Nature*, 466 (7302): 29.
- Kagan, S. (2001). Thinking about Cases, *Social Philosophy and Policy*, 18 (2): 44–63.
- Klein, E., S. Goering, J. Gagne, C. V. Shea, R. Franklin, S. Zorowitz, ... A. S. Widge (2016). Brain-Computer Interface-Based Control of Closed-Loop Brain Stimulation: Attitudes and Ethical Considerations, *Brain-Computer Interfaces*, 3 (3): 140–148.
- Klenk, M. (2020). Charting Moral Psychology’s Significance for Bioethics: Routes to Bioethical Progress, its Limits, and Lessons from Moral Philosophy, *Diametros*, 17 (64): 36–55.
- Knobe, J. (2007). Experimental Philosophy, *Philosophy Compass*, 2 (1): 81–92.
- Knobe, J. (2016). “Experimental Philosophy is Cognitive Science,” in J. Sytsma and W. Buckwalter (eds.), *A Companion to Experimental Philosophy*, 37–52, Oxford: Wiley-Blackwell.
- Knobe, J. (2019a). Philosophical Intuitions are Surprisingly Robust Across Demographic Differences, *Epistemology & Philosophy of Science*, 56 (2): 29–36.
- Knobe, J. (2019b). Difference and Robustness in the Patterns of Philosophical Intuition across Demographic Groups, unpublished manuscript. Available online: <https://cpb-us-w2.wpmucdn.com/campuspress.yale.edu/dist/3/1454/files/2019/12/Difference-Robustness-2.pdf> (accessed October 26, 2021).
- Knobe, J. (in press). “Personal Identity and Dual Character,” in K. P. Tobia (ed.), *Experimental Philosophy of Identity and the Self*, London: Bloomsbury.
- Knobe, J., W. Buckwalter, S. Nichols, P. Robbins, H. Sarkissian, and T. Sommers (2012). Experimental Philosophy, *Annual Review of Psychology*, 63: 81–99.

- Kuorikoski, J. and C. Marchionni (2016). Evidential Diversity and the Triangulation of Phenomena, *Philosophy of Science*, 83 (2): 227–247.
- Lewis, J. (2020a). From X-phi to Bioxphi: Lessons in Conceptual Analysis 2.0, *AJOB Empirical Bioethics*, 11 (1): 34–36.
- Lewis, J. (2020b). Getting Obligations Right: Autonomy and Shared Decision Making, *Journal of Applied Philosophy*, 37 (1): 118–140.
- Lewis, J. (2021). Autonomy and the Limits of Cognitive Enhancement, *Bioethics*, 35 (1): 15–22.
- Locke, J. (1689/1694). *An Essay Concerning Human Understanding*, P. H. Nidditch (ed.), Oxford: Oxford University Press, 1975.
- Machery, E. (2017). *Philosophy Within its Proper Bounds*, Oxford: Oxford University Press.
- Machery, E. and S. Stich (2019). Demographic Differences in Philosophical Intuition: A Reply to Joshua Knobe, unpublished manuscript. Available online: <https://www.dropbox.com/s/flndq44xjtyrsyi/Reply%20to%20Knobe%20--%20Demographic%20Difference%20in%20Philosophical%20Intuition%20-%2011-3-2019.docx?dl=0> (accessed October 26, 2021).
- Mackenzie, C. and N. Stoljar (2000). “Autonomy Refigured,” in C. Mackenzie and N. Stoljar (eds.), *Relational autonomy: Feminist Perspectives on Autonomy, Agency and the Social Self*, 3–31, Oxford: Oxford University Press.
- McMillan, J. (2018). *The methods of bioethics: an essay in meta-bioethics*. Oxford: Oxford University Press.
- Meyers, D. T. (2005). “Decentralizing Autonomy: Five Faces of Selfhood,” in J. Christman and J. Anderson (eds.), *Autonomy and the Challenges to Liberalism: New Essays*, 27–55, Cambridge: Cambridge University Press.

- Mihailov, E., I. R. Hannikainen, and B. D. Earp (2021). Advancing Methods in Empirical Bioethics: Bioxphi Meets Digital Technologies, *The American Journal of Bioethics*, 21 (6): 53–56.
- Mukerji, N. (2019). *Experimental Philosophy: A Critical Study*, London: Rowman & Littlefield.
- Nado, J. (2021). Conceptual Engineering Via Experimental Philosophy, *Inquiry*, 64 (1–2): 76–96.
- Neiders, I., and V. Dranseika (2020). Minds, Brains, and Hearts: an Empirical Study on Pluralism Concerning Death Determination, *Monash Bioethics Review*, 38 (1): 35–48.
- Nichols, S. and M. Bruno (2010). Intuitions about Personal Identity: An Empirical Study, *Philosophical Psychology*, 23 (3): 293–312.
- O’Neill, E. and E. Machery (2014). “Experimental Philosophy: What is it Good For?,” in E. Machery and E. O’Neill (eds.), *Current Controversies in Experimental Philosophy*, vii–xxix, London: Routledge.
- Persson, I. and J. Savulescu (2012). *Unfit for the Future: The Need for Moral Enhancement*, Oxford: Oxford University Press.
- Prinz, J. J. and S. Nichols (2016). “Diachronic Identity and the Moral Self,” in J. Kiverstein (ed.), *The Routledge Handbook of Philosophy of the Social Mind*, 465–480, London: Routledge.
- Reid, T. 1785. *Essays on the Intellectual Powers of Man*, Cambridge, MA: MIT Press, 1969.
- Reiner, P. B. (2019). “Experimental Neuroethics,” in S. K. Nagel (ed.), *Shaping Children: Ethical and Social Questions That Arise When Enhancing the Young*, 75–83, Cham: Springer.

- Schlegel, R. J., J. A. Hicks, J. Arndt, and L. A. King (2009). Thine Own Self: True Self-Concept Accessibility and Meaning in Life, *Journal of Personality and Social Psychology*, 96 (2): 473–490.
- Schlegel, R. J., J. A. Hicks, L. A. King, and J. Arndt (2011). Feeling Like You Know Who You Are: Perceived True Self-Knowledge and Meaning in Life, *Personality & Social Psychology Bulletin*, 37 (6): 745–756.
- Schlegel, R. J. and J. A. Hicks (2011). The True Self and Psychological Health: Emerging Evidence and Future Directions, *Social and Personality Psychology Compass*, 5 (12): 989–1003.
- Schüpbach, M., M. Gargiulo, M. L. Welter, L. Mallet, C. Béhar, J. L. Houeto, ... Y. Agid (2006). Neurosurgery in Parkinson disease: a Distressed Mind in a Repaired Body?. *Neurology*, 66 (12): 1811–1816.
- Shoemaker, D. (2010). The Insignificance of Personal Identity for Bioethics, *Bioethics*, 24 (9): 481–489.
- Skorburg, J. A., and W. Sinnott Armstrong (2020). “Some Ethics of Deep Brain Stimulation,” in D. Stein and I. Singh (eds.), *Global Mental Health and Neuroethics*, 117–132, London: Academic Press.
- Strohming, N. and S. Nichols (2014). The Essential Moral Self, *Cognition*, 131 (1): 159–171.
- Strohming, N. and S. Nichols (2015). Neurodegeneration and Identity, *Psychological Science*, 26 (9): 1469–1479.
- Thomson, J. J. (1986). *Rights, Restitution, and Risk*, Cambridge, MA: Harvard University Press.
- Tobia, K. P. (2015). Personal Identity and the Phineas Gage Effect, *Analysis*, 75 (3): 396–405.

- Tobia, K. P. (2016). Personal Identity, Direction of Change, and Neuroethics, *Neuroethics*, 9 (1): 37–43.
- Tobia, K. P. (2017). Change Becomes You, *Aeon*. Available online: <https://aeon.co/essays/to-be-true-to-ones-self-means-changing-to-become-that-self> (accessed October 26, 2021).
- Veatch, R. M. and L. F. Ross (2016). *Defining Death: The Case for Choice*, Washington, DC: Georgetown University Press.
- Veit, W., B. D. Earp, H. Browning, and J. Savulescu (in press). Evaluating Trade-Offs Between Autonomy and Well-Being in Supported Decision Making, *American Journal of Bioethics*.
- Wedgwood, R. (2007). *The Nature of Normativity*, Oxford: Oxford University Press.
- Wilkes, K. V. (1988). *Real People: Personal Identity Without Thought Experiments*, Oxford: Oxford University Press.
- Williams, B. (1970). The Self and the Future, *The Philosophical Review*, 79 (2): 161–180.
- Witt, K., J. Kuhn, L. Timmermann, M. Zurowski, and C. Woopen (2013). Deep Brain Stimulation and the Search for Identity, *Neuroethics*, 6 (3): 49–511.