The Whole Story: Identity and Narrative

Marya Schechtman
University of Illinois, Chicago

Abstract:

The burgeoning use of experimental methods to consider questions of human nature and personal identity has been a fruitful and exciting development, yielding significant and provocative results. This essay argues for the value of including reflection on the treatment of these topics in fictional narratives to complement and deepen results in experimental philosophy. Experimental vignettes are by necessity brief and schematic. This is part of what makes them so effective in the experimental context. The space afforded for detail, complexity, and ambiguity by the format of fiction allows it to highlight and explore issues that cannot easily be incorporated into experimental method. By juxtaposing a fictional narrative in which we are led to view a character as fundamentally bad with a structurally similar experimental vignette in which participants judge the protagonist to be fundamentally good, I demonstrate how reflection on fiction can contribute to debates in experimental philosophy and reveal distinctions between different dimensions of questions about identity and morality.

Keywords: personal identity, morality, human nature, fiction, narrative

The burgeoning use of experimental methods to consider questions of human nature and personal identity has been a fruitful and exciting development. Rigorously designed studies investigating how we think about these concepts have yielded robust and sometimes surprising results, especially concerning our moral evaluation of the true self. As welcome and important as this new methodology has proved, this essay argues for the value of focusing also on another long-used means of gaining insight into how we think about identity and human nature, namely reflection on the treatment of these themes in narrative fiction. The case for fiction is not based on a suggestion that experimental methods are defective, only incomplete. The brevity and clarity of experimental vignettes, which is a great strength of the experimental approach, also imposes limitations. The format of fiction allows for exploration of forms of messiness and ambiguity inherent in the human condition that are difficult to investigate in an experimental context. Questions of identity and morality are complex and multidimensional, and different aspects of these questions call for different approaches. Bringing these two approaches into dialogue thus promises to yield broader and richer insights than are obtained by relying on either alone.

I begin with a brief review of some experimental results that have been interpreted as implying that we tend to believe people are fundamentally morally good (section 1), after which I look at an alternate interpretation of these results that has been offered within experimental philosophy (section 2). I turn next to an example from fiction which touches on the debate introduced in sections 1 and 2, reflecting on the way in which viewers are led to judge Tony Soprano in the television series *The Sopranos* to be intrinsically bad (section 3). Next, I contrast how viewers make moral judgments about Tony Soprano with how participants make moral judgments in responding to an experimental vignette involving a professional assassin. This juxtaposition shows how the more expansive nature of fiction can explain the divergent reactions fond in these two contexts, underscoring the importance of remaining clear on the distinction between generic and individual questions about moral nature (section 4).

1. The True and the Good

A striking series of recent studies suggests that we tend to believe that people are fundamentally good "deep down". This conclusion is gleaned from a variety of experiments which reveal asymmetrical responses to a range of different vignettes depending upon their moral valence. For instance, when an agent acts on desires that conflict with her beliefs participants are more inclined to say that she "values" what she is doing if they view her desires as morally good than if they view them as morally objectionable (Knobe & Roedder, 2009). Someone who acts on overwhelming and irresistible emotion is judged by participants to deserve less blame for a bad action than someone who acts coolly on considered motives, even though someone who is driven by overwhelming emotion to perform a morally good action is not typically seen as less praiseworthy than someone who undertakes such an action as the result of coolheaded deliberation but, if anything, as more praiseworthy. (Pizarro, et al., 2003)

One explanation that has been given for these and related asymmetries in judgement is that they stem from the implicit assumption that the true self is morally good. Faraci and Shoemaker (2019, p. 606) call this the "Good True Self" (GTS) theory. Further investigation, aimed at testing directly whether assumptions about the nature of the true self mediate these responses, supports this interpretation. Newman, De Freitas, and Knobe (2015), for instance, undertake a series of five experiments. Four reproduce existing experiments that generate asymmetrical judgments based on moral valence with the addition of a separate question about whether the relevant actions or attitudes represent the agent's true self. The fifth directly manipulates beliefs about the true self and measures effects on the application of the concepts involved in the asymmetries.

For what follows it will be useful to have an example in hand, and here I choose one that will serve as a useful foil to the fictional case I will consider later. One of the studies Newman et. al. undertook replicated an earlier study by Sousa & Mauro (2015) involving asymmetries in judgments concerning weakness of will. Sousa & Mauro gave one group of participants the following vignette:

> John is a professional assassin. He has started to think about quitting this profession because he feels that it is wrong to kill another person. However, he is strongly inclined to continue with it because of the financial benefits.

> John is in conflict, but after considering all aspects of the matter, he concludes that the best thing for him to do is to quit his profession. Accordingly, he decides that the next day he will look for a job that does not involve violence.
>
> The next day, while still completely sure that the best thing for him to do is to look for a job that does not involve violence, John is swayed by the financial benefits. Against what he had decided, he kills another person for money.

Another group received a modified version of this vignette:

> John is in conflict, but after considering all aspects of the matter, he concludes that the best thing for him to do is to continue with his profession. Accordingly, he decides that the next day he will kill another person for money.
>
> The next day, while still completely sure that the best thing for him to do is to kill another person for money, John is swayed by the feeling that it is wrong to kill. Against what he had decided, he looks for a job that does not involve violence. (Reproduced in (Newman, et al., 2015, pp. 14-16))

Although each condition involves John acting against what he has decided to do, and so, on most philosophical accounts, as exhibiting weakness of will, participants tended to judge him to be weak of will only in the first condition, where he acts immorally against his good impulses, and not in the second, where he acts morally against immoral impulses.

Newman et. al. replicated these results, but also asked participants to respond to a question about whether the agent went against his true self in the action. The results indicated that beliefs about the true self did in fact mediate the judgments about weakness of will. The other studies produced similar results. These findings are taken to provide strong evidence for the view that the asymmetries based on moral valence are, as the authors suggest, “symptoms of a single unified phenomenon: the tendency to assume that, deep down, others are morally good." (Newman, et al., 2015, p. 28). Other studies e.g., (Tobia, 2015) have produced similar results. There is, however, always room for another point of view. The next section considers an alternate interpretation of these data proposed by Faraci and Shoemaker.

2. An Alternative Explanation

Faraci and Shoemaker approach these issues through work on attributability theory, the view, roughly, that "for an agent to be the proper target of praise or blame *for* the action of a particular moment, that action must be expressive *of* that agent" (2019, p. 607), it must, in other words, represent the deep or true self. In this context, Faraci and Shoemaker (2014) had conducted earlier studies that had revealed asymmetries in attributions of praise and blame consistent with those found by others. These studies employed four scenarios involving a white male named Tom. Tom$_A$ was raised in New Orleans and taught to respect all people equally but as an adult decides to embrace the identity of a proud racist. When he is 25, he encounters a black man who has tripped and fallen outside his home and, in keeping with his beliefs, spits on the man as he walks by. Tom$_B$ was raised on an isolated island in the bayous of Louisiana and was taught, growing up, that all non-white people are inferior. As an adult he embraces what he has been taught. When he is 25, he encounters a black man who has tripped and fallen outside his home and, in keeping with his beliefs, spits on the man as he walks by. Participants overall found Tom$_A$ to be more blameworthy for his action than Tom$_B$. The presumption was that Tom$_B$'s morally impoverished background led subjects to attribute his action to him less fully, and hence to find him less culpable, than Tom$_A$.

At the same time, however, they included a second set of cases about which they surveyed additional participants. Tom$_C$, like Tom$_A$, was taught to respect all people equally but decided, as an adult, to embrace racism. Like the previous Toms, at age 25 he encounters a black man who has fallen in front of his home, but unlike them, Tom$_C$ goes against his considered moral beliefs and helps the man up. Tom$_D$, like Tom$_B$, was raised on an isolated island and taught to believe that all non-white people are inferior, a view he embraces as an adult. Like the others, he encounters a fallen black man in front of his house and, like Tom$_C$ helps him up despite his beliefs that he should spit on him. In these cases, subjects were asked about praiseworthiness, and were inclined to say that Tom$_D$ was *more* praiseworthy than Tom$_C$. This asymmetry raises questions about the simple hypothesis that a morally deprived upbringing decreases attributability and so mitigates both blame and praise.

Faraci and Shoemaker's original thought was that this difference could be accounted for by the "Difficulty Hypothesis", that moral ignorance of the sort depicted in the cases of Toms B and D affects assessments of blameworthiness or praiseworthiness insofar as it affects the level of difficulty involved in doing the right thing. The idea is that it is more difficult for Tom$_B$ than for Tom$_A$ to help the black man, given how he was raised, so he is less blameworthy. Likewise, it was more difficult for Tom$_D$ than for Tom$_C$ to recognize the rightness of helping him up, so he deserves extra praise. While this is an intuitive explanation for the asymmetries, the work of Newman et. al, inspired them to undertake a new study to see if the response was mediated by beliefs about the true self. Their results replicated those of Newman et. al., suggesting that despite their parallel childhood deprivation and moral ignorance, participants judged Tom$_A$ and Tom$_B$ as both being less themselves when spitting on the black man, and Tom$_C$ and Tom$_D$ as both being more themselves when they helped him up.

While these findings do support the GTS theory, Faraci and Shoemaker point out that these are not the only relevant data. Within these and other experiments are also examples of cases where we judge bad actions to be more representative of the true self than good ones. In their own studies, they note, subjects tended to judge Tom$_A$ to be more blameworthy for spitting on the black man than Tom$_C$ is seen to be praiseworthy for helping him up. If the true self were straightforwardly figured as the good self, one would expect the opposite to be true. Earlier work by Knobe, moreover, showed that when presented with a case where a CEO doesn't care about how a decision about company policy will predicably affect the environment, participants tend to judge those environmental effects as intentional when they are detrimental, but unintentional when they are positive. ((Knobe, 206) quoted in (Faraci & Shoemaker, 2019, p. 616))

Faraci and Shoemaker (2019, p. 619) acknowledge that these data do not show the GTS theory to be false, but they do speak against the claim that it is the obvious interpretation of experimental results. They thus suggest an alternative hypothesis. This alternative interprets our patterns of attribution as showing that we judge bad behavior as representative of the true self in the absence of another available explanation. It is widely recognized, they point out, that imperfect conditions can cause someone who is not intrinsically bad to act badly. When someone behaves badly, it *might* be because she is a bad person deep down, but it also might be because she is under intense stress or overcome with emotion she cannot control. As an alternative to the GTS theory, which sees the data as showing that participants assume that people are good deep down, and so that bad actions almost never represent the true self, Fraraci and Shoemaker propose instead what I will call the "benefit-of-the-doubt" (BD) theory. BD theory holds that that when an excusing explanation of bad behavior is available we will, to a point at least, give the agent the benefit of the doubt, assuming that it is some interfering condition rather than intrinsic badness that is responsible for the bad action. (2019, pp. 617-620)

Faraci and Shoemaker acknowledge that this is only one possible interpretation of the data, and that the GTS theory can also explain them. They thus suggest the need for further investigation of the precise nature of our moral assessments in these cases and their relation to judgments about the true self. Presumably, they are thinking most immediately of further work in experimental philosophy, and undoubtedly there is a great deal of progress to be made in this way. There is, however, another venue in which questions about our patterns of attribution and assessment are explored, and that is narrative fiction. There is reason to hope that such explorations also have something to add to our overall attempts to understand these difficult matters. In the next section I describe a narrative that bears directly on the questions Faraci and Shoemaker raise.

3. The Good, The Bad, and the Complicated

The past few decades have been hailed by many as a golden era of television because of the rise of "prestige dramas" like *Breaking Bad,* and *The Sopranos,* which are recognized as serious aesthetic accomplishments. In each of these dramas we are presented with a complex, charismatic, and fascinating character who engages in horrific, immoral behavior. It is generally agreed that viewers who watch these series are ultimately forced to acknowledge that these characters are truly bad deep down.

The protagonists of these series are what A. W. Eaton calls, taking a phrase from Hume, "rough heroes". (2012, p. 281) The defining features of rough heroes can be seen by contrasting them with antiheroes. The latter, though flawed, have redeeming features which ultimately outweigh their deficits. The rough hero, by contrast, has flaws that are “grievous” and “an integral part of his personality rather than peripheral failings or foibles”. He “fully intends to do bad and is remorseless about his crimes”. The narratives in which these heroes are portrayed do not offer “reasons to dismiss his misdeeds as the result [of] misfortune, weakness, folly, or ignorance”, and his “vices are not outweighed by some more redeeming virtues.” (Eaton, 2012, p. 284) Examples of rough heroes outside of prestige dramas include Humbert Humbert in *Lolita*, Satan in *Paradise Lost,* and Alex in *A Clockwork Orange*. It is a critical feature of rough heroes, Eaton argues, and the unique aesthetic accomplishment of the works that contain them, that we simultaneously recognize them as deeply and intrinsically bad and feel a sympathetic attraction to them. To get the viewer to see them as they truly are, it is necessary to get them to overcome whatever tendencies we have to see people as fundamentally good or benefit of the doubt we are inclined to give in judging them. By offering an extended and detailed picture of a case in which we do overcome initial impulses to judge a character positively, these dramas thus potentially have something to say about when and how we come to judge someone to be fundamentally bad, including the role excusing explanations might play.

To explore what might be learned by reflection on a fictional narrative, I take Tony Soprano as my example, since his case will present a useful contrast with the vignette about John the assassin presented earlier. There is general critical agreement that Tony Soprano is an extraordinarily complex and intriguing character. He is a mob boss, who lives an extremely violent life and regularly commits monstrous, immoral acts. He is also, however, a husband and father who loves his family and friends and does a great deal to enhance their wellbeing. He is charming, affable, and intelligent. We learn that Tony had an upbring of deep moral deprivation, that his mother was an abusive and uncaring parent, and that growing up he received none of the love or attention humans need to thrive. As an adult, he is vulnerable and sensitive. At the beginning of the series, he suffers a panic attack which leads him to therapy. His therapist links his symptoms to the departure of a family of ducks that had been nesting in his pool. His investment in the ducks, she suggests, represents his deep investment in his family and fear of abandonment. Throughout the series we see him attend therapy sessions and undertake various other efforts aimed at self-improvement, all of this alongside intensely vicious actions and general moral depravity.

We have, of course, seen complicated and conflicted mob bosses before in fiction. The mafioso with a heart of gold, who commits despicable acts but is fiercely loving and loyal to family and friends, has a rigid code of honor that he follows scrupulously, and is pushed to his life of crime through external forces, is a paradigmatic antihero. What makes Tony Soprano different is that although the series frequently invites the viewer to understand him in just this way, it also repeatedly undercuts attempts to do so. He *is* genuinely good to his family and friends, often supporting and mentoring them, but he also regularly betrays and hurts them, sometimes turns on them violently, and is willing to murder those close to him if they become inconvenient, seemly with no real remorse. He *is* vulnerable and sensitive at times, but he is also often remarkably callous and unconcerned about the pain he causes others, even those he loves. He gestures toward a code of honor, but violates it readily, and although he does try to improve, he is not interested in doing any difficult work in this regard and easily becomes bored with the project. Tony is without question an immensely complicated person. Unlike the antihero mafioso, however, in the end, he does not seem terribly conflicted.

Crucial for our purposes is the way in which the different sides of Tony's personality emerge over time. The viewer starts out clear that Tony is a violent man who acts immorally but is also immediately exposed to his positive traits and vulnerabilities. Moving the viewer from an assessment of Tony as a tragic antihero to the recognition that he is a truly disturbing and violent rough hero, takes a carefully orchestrated set of incidents through which viewers are repeatedly invited to accept an excusing explanation of Tony's behavior of the sort Faraci and Shoemaker suggest, only to see him then behave in a way so violent or insensitive or malicious that it is not only difficult to find any explanation other than malevolence for the current behavior but becomes increasingly difficult to believe that bad behavior explained away earlier is not, after all, representative of who Tony truly is deep down. After many iterations of this cycle, the viewer is forced to conclude that Tony is, as Eaton puts it, "a liar, a thief, an extortionist, and a womanizer; he is pathologically callous, selfish, bigoted, racist, homophobic, and self-centered." (2012, p. 281)

I have suggested that reflection on fictional treatments of these issues can provide useful insights to complement experimental work. In the next section I provide an example of the way in which thinking about *The Sopranos* might contribute to discussion of these topics.

4. Implications

One way to see the kind of contribution fiction is especially well-placed to make is to contrast the story of Tony Soprano with the vignette of John the assassin, mentioned earlier. A television series that follows a character over multiple seasons will obviously contain a great deal more detail about his behavior and the possible causes for it than a vignette designed for experimental purposes. In the case of Tony and John, this is a difference that potentially makes a difference in the way we assess their behavior and its relation to their true nature. There are isolated incidents within the trajectory of Tony Soprano in which he decides not to carry out some particular act of violence but ends up doing so anyway. Presented with only this sequence of events, the viewer is likely to judge that Tony has displayed weakness of will as participants in the experimental studies judge John to do. This would be the initial step in the dynamic described in the last section. Seen within the context of the entire series, however, viewers are likely to ultimately revise that initial judgement and conclude that in fact Tony did exactly what he wanted to do. That is the final step of an extended dynamic that is possible to portray in fiction, but not in an experimental vignette.

What, then, do we learn from considering the more extended and detailed depiction of Tony's reversal rather than the more abbreviated one that we find in the case of John? One might take the differing judgments in the cases of John and Tony to support BD theory over the GTS theory. It seems that in viewing *The Sopranos* viewers do offer Tony the benefit of the doubt up to a point, but when excusing explanations are no longer viable, they withdraw it and conclude that he is bad. While I find this an intriguing possibility, it is a conclusion that we cannot draw without more work. There are two reasons. First, the GTS theory does allow that we can, in rare and extreme circumstances, judge someone to be fundamentally bad. The data, after all, do not show that all participants always judge everyone to be entirely good. Second the case of Tony does not quite show excusing explanations to be discarded. While ultimately viewers tend to judge that he cannot be excused for what he does, it is not as if what had been taken as excusing explanations are shown to be false; they just play a somewhat different role in his life than we originally thought. Further reflection on the nature of excusing explanations thus seems indicated. While I think that there is real promise that further reflection on Tony's case can shed light on the dispute between the GTS theory and BD, more work is required to determine just what the implications are.

Another possible conclusion is that the judgments that come from reflection on fictional accounts is somehow more reliable than that found in experimental philosophy, since it is based on more complete information. There is clearly a great deal that speaks against this conclusion, however, at least as a general claim. The detail found in fictional narrative is bought at the cost of the control, quantifiability, reproducibility, and generalizability experimental philosophy offers. *The Sopranos* generates judgements about and reactions to one (fictional) man, one that leads a very unusual life. There is no straightforward route from our judgments about Tony Soprano to our judgments about people in general. I have talked about how viewers react to Tony's exploits, but data about how "we" react to Tony, while I believe it to be generally accurate, is collected in a highly unsystematic way. There is nothing, in the case of fiction, like the concrete responses to survey questions found in experimental philosophy, and so no firm data about what percentage of viewers felt exactly what way about Tony's actions or true self. In multiple senses, then, the judgments that come from the data in experimental philosophy are more reliable than those that come out of discussing fiction.

The contributions of reflection on fictional narratives that I would argue for thus lie not in providing additional or different data of the sort that experimental philosophy collects. Fiction will always be worse at that. Juxtaposing these different ways of thinking about moral assessment of the true self can, however, help to distinguish different kinds of questions about the moral assessment of the true self, and to see that different methods may be indicated depending on which we are asking. Looking at the cases of Tony and John side-by-side uncovers a certain ambiguity about what question is being answered by participants in the experimental studies. On the one hand, there is a question about human nature – whether humans as a kind are fundamentally good. On the other, there is a question about a particular person, whether Phineas, or John, or Tony is fundamentally good.

The experimental vignettes, in some sense, ask about particular people. There are protagonists, and they have names. Since the information given about each is so limited, however, it is not clear that they are really being interpreted as particular individuals by the participants as opposed to being seen as generic representatives of humankind. In everyday life, it is a truism that if you want to know whether someone is truly good (or bad) deep down, you need to know more than a few sentences about his history and the circumstances of one particular action he took. Nonetheless, participants in the experimental setting subjects *do* make such judgements, based on just such information, and the judgments they make are remarkably consistent. Since there is little particular information about the people in these vignettes, and since it is, after all, general information we are looking for, it is reasonable to assume that in making them participants are consulting their generic views about what humans are like (as the GTS theory implies) or what protocols we should use for judging humans in general (as BD theory suggests). Either way, however, these results yield conclusion abouts how we morally assess *people* in general, and not about how we assess a particular person.

The difference between these questions can be made sharper by looking at how experimental philosophers have characterized what it means to say that we take the true self to be fundamentally good. Shoemaker and Tobia (Forthcoming) mention two possibilities for thinking about the picture of the true self that explains the asymmetries in moral assessment found in the experimental data. One refers to essences, suggesting that human nature is seen as essentially good and so that any deviation from goodness is a corruption of one's true essence, e.g., (Strohminger, et al., 2017) Another emphasizes teleological persistence, suggesting that moral improvement represents a person's true purpose, e.g., (Rose, et al., 2020) What these and related accounts have in common is that claims about the good true self are normative rather than descriptive, either insofar as fundamental goodness is seen as criterial for humanness or as the proper purpose of humans. Neither version implies that all, or even most, individual humans are actually good, only that insofar as they are not good, they deviate from the way humans are "meant" to be.

The question of individual nature, however, while it is a moral judgment and so normative in that sense, is descriptive insofar as it captures the particularities that make an individual the unique character that he is, including the various ways in which he deviates from generic human nature. To say that Tony is fundamentally bad is say that his badness is an intrinsic part of who he is as an individual. Determining this, I have suggested, requires more information than can be supplied in an experimental vignette. If we are interested in determining whether *this* person is fundamentally good or evil, rather than whether *people* are, it is necessary to look at the confluence of choices and influences and opportunities in his life. Presenting these details is something that fiction is well suited to do. Following an individual through a narrative and coming to an ultimate judgment about his nature will not provide a set of general conditions under which we judge an individual to be good or bad but, as I hope the case of Tony Soprano shows, it can provide general insights about what goes into making such determinations.

Recent work by Knobe (Forthcoming) challenges the distinction I have just drawn between questions about the fundamental nature of generic humans and that of individuals. Knobe proposes using the framework of dual character concepts to explain asymmetric judgments in cases of moral change. This framework applies in situations where we recognize that an individual formally falls under a particular category, but also feel that it fails to instantiate the truest norms of that category. The person who engages in scientific research but has no intellectual curiosity is in some sense a scientist but is "when you think about it" not really a scientist at all. Similar claims may be made of the artist who writes and performs punk rock but is motivated by banal commercialism, or the churchgoer who is merely going through the motions. Knobe suggests that something similar may be at work in our assessments of identity through moral change.

On the surface, this framework seems to favor the distinction between generic and individual questions of one's nature that I have been urging. We have on the one hand the question of what it means to *really* be a scientist, or Christian, or punk rocker, and on the other a question about *this* individual who in some crude, formal sense falls under one of these categories but fails to instantiate the true nature of the kind. Knobe argues, however, that preliminary research suggests that we make similar kinds of judgments about individuals. When subjects are presented with a scenario like that in Tobia (2015) where someone named Phineas changes from a morally good person to a morally bad one, they tend to agree with the statement: "There's a sense in which the man after the accident is clearly still Phineas, but ultimately, if you think about what it really means to be Phineas, you'd have to say that he is not truly Phineas at all." (Knobe, Forthcoming) This can be interpreted to show that the nature of the individual is just an instance of the nature of the kind. Since Phineas is a human, what it means to "really be" Phineas is what it "really means" to be human, and when Phineas loses the good essence of humankind, he ceases to be himself.

If this is right, the question I suggested was better addressed by fiction just *is* an example of the general question experimental philosophy is especially well-placed to answer. This is not, however, the only way to interpret Knobe's results. Given the very limited information provided about Phineas in these vignettes, participants may well be taking him to stand in for generic humanity in the way I described above. Phineas Gage is a specific individual, but this Phineas is really just “a person named Phineas”. Some support for this suggestion is found in the results obtained in the condition where Phineas is described as changing from a morally bad person to a morally good one. When participants in this condition are presented with the question about whether he is truly Phineas at all after the change, their “judgments were all over the place. Some agreed with the dual character statement, others disagreed.” (Knobe, Forthcoming) A possible explanation for this result is that in the moral improvement case, Phineas’ initial moral badness presents him immediately as an individual who does *not* instantiate paradigmatic humanness, and this may signal some participants to interpret this as a case in which they are asked about the individual nature of a specific person rather than about generic humanity. The variable answers may thus result from the difficulty of making such judgments with limited information.

Obviously, more investigation is needed to draw any conclusions here, and additional work in experimental philosophy is going to be an essential part of this investigation. If the distinction between generic and individual questions can be maintained, however, there is also work which reflection on fictional treatments of these matters is especially well suited to do. In any event, bringing these considerations into the mix can only help us to recognize that these questions can be asked either about particular individuals or about humans as a kind, and to think more clearly about how these two types of questions are related to one another.

Experimental work is without question a valuable addition to philosophical discussions of personal identity. It provides controlled, qualified, and reproducible data about how key concepts are understood, and has produced a great many valuable and provocative results. The very strengths of these experiments in answering the questions for which they were designed can, however, represent drawbacks in other contexts. There are many dimensions to our thinking about human nature, and correspondingly different questions about the moral nature of the true self. To map out, let alone civilize, this messy and exciting terrain will require multiple tools in our methodological toolbox. Luckily, we have a great many, and there is a great deal to be hoped for as we learn how to use them together.[1]

[1] I am grateful to Kevin Tobia, Joshua Knobe, and David Shoemaker for immensely helpful comments on an earlier draft of this paper.

Bibliography

Eaton, A., 2012. Robust Immoralism. *The Journal of Aesthetics and Art Criticism,* 70(3), pp. 281-292.

Faraci, D. & David, S., 2014. Huck vs. JoJo: Moral Ignorance and the (A)symmetry of Praise and Blame. In: T. Lombrozo, J. Knobe & S. Nichols, eds. *Oxford Studies in Experimental Philosophy.* Oxford: Oxford University Press, pp. 7-27.

Faraci, D. & Shoemaker, D., 2019. Good Selves, True Selves: Moral Ignorance, Responsibility, and the Presumption of Goodness. *Philosophy and Phenomenological Reserach,* 98(3), pp. 606-622.

Knobe, J., 206. The Concept of Intentional Action: A Case Study in the Uses of Folk Psychology. *Philosophical Studies,* Volume 130, pp. 203-31.

Knobe, J., Forthcoming. Personal Identity and Dual Character Concepts. In: K. Tobia, ed. *Experimental Philosophy of Identity and the Self.* s.l.:Bloomsbury.

Knobe, J. & Roedder, E., 2009. The ordinary concept of valuing. *Philosophical Issues,* Volume 19, pp. 131-47.

Locke, J., 1975. *An Essay Concerning Human Understanding.* Oxford: Oxford University Press.

Newman, G. E., De Freitas, J. & Knobe, J., 2015. Beliefs About the True Self Explain Asymmetries Based on Moral Judgement. *Cognitive Science ,* Volume 39, pp. 96-125.

Pizarro, D. A., Uhlmann, E. & Salovey, P., 2003. Asymmetry in judgments of moral blame and praise: The role of perceived metadesires. *Psychological Science,* Volume 14, pp. 267-72.

Rose, D., Tobia, K. & Schaffer, J., 2020. Folk teleology drives persistence judgments. *Synthese,* Volume 197, pp. 5491-5509.

Shoemaker, D. & Tobia, K., Forthcoming. Personal Identity. In: M. Varga & J. Doris, eds. *Oxford Handbook of Moral Psychology.* Oxford: Oxford University Press.

Sousa, P. & Mauro, C., 2015. The evaluative nature of the folk concepts of weakness and strength of will. *Philosophical Psychology,* 28(4), pp. 487-509.

Strohminger, N., Knobe, J. & Newman, G., 2017. The true self: A psychological concept distinct from the self. *Perspectives on Psychological Science,* Volume 12, pp. 551-560.

Tobia, K. P., 2015. Personal identity and the Phineas Gage effect. *Analysis,* Volume 75, pp. 396-405.