



# HOKKAIDO UNIVERSITY

|                  |                                                                                     |
|------------------|-------------------------------------------------------------------------------------|
| Title            | A Realist Self?                                                                     |
| Author(s)        | Oberg, Andrew                                                                       |
| Citation         | Journal of Applied Ethics and Philosophy, 7, 24-33                                  |
| Issue Date       | 2015-09                                                                             |
| DOI              | <a href="https://doi.org/10.14943/jaep.7.24">https://doi.org/10.14943/jaep.7.24</a> |
| Doc URL          | <a href="https://hdl.handle.net/2115/68222">https://hdl.handle.net/2115/68222</a>   |
| Type             | departmental bulletin paper                                                         |
| File Information | JAEP7_24-33.pdf                                                                     |



# A Realist Self?

Andrew Oberg

Toyo University

## Abstract

Since the demise of the Cartesian dualist view of the self a number of possible definitions of what the self could be, if indeed it can be said to be anything, have been put forward but no consensus has yet been reached. In fact, such seems a long way off. In what follows four accounts of the self that are representative of the broad trends in the literature are analyzed for theoretical vigor and empirical accuracy in light of recent advances in cognitive studies and the findings of psychological research into behavior and decision-making. The self-concepts examined are of both the anti-realist and realist varieties, with one particular realist account found to be most apposite. The account is not without its flaws, however, and as such an alternative self view is offered that builds on and adds to its strengths. Finally, some ethical implications of adopting the proffered self-concept are considered.

Key words: anti-realist self; Hume; Kristjánsson; nonself; realist self

## 1. Outline of the Study

The search for an accurate description of the self has been a centuries long affair, and the following will not attempt to definitively solve this problematic. Rather, this paper will instead focus on exploring four common theoretical accounts of the self, examining each for philosophical robustness and correspondence to recent research in cognitive and psychological studies. The four accounts to be considered have been chosen for their representational qualities and are not meant to be an exhaustive analysis of the work done on the self heretofore. Largely following Kristjánsson's descriptive categories in *The Self and Its Emotions* (Kristjánsson 2010), the first account will consider the soft anti-realist position, the second will cover the hard anti-realist position, and the third the soft realist position. Our final account will be of the contextualized (or conditioned) soft realist position, a category that Kristjánsson did not consider and whose title I have had to create. The hard realist position, that of a separately existing and (usually) eternal Cartesian ego or soul comprising the self and temporarily housed in the body, will not be examined due to its tendency to preempt further debate on the topic and

its general rejection by the academic community save for historical or religious discussions. Our study will then conclude with a proposal for an alternative version of Kristjánsson's self-concept which would be more in line with the interdisciplinary research presented and could offer important ethical benefits were such a view to be adopted.

## 2. Four Accounts of the Self

### *A. The soft anti-realist position, or The (non-communal) libertarian atomistic self*

This is the account of the self that most of us hold without really thinking about it and upon which the majority (if not entirety) of present Western-style legal charters are based, and so it is from this angle – the political – that we are best able to tease out what it entails since the position is generally taken for granted and not explicitly argued for. Legal structures such as bills of rights bestow certain privileges and offer certain protections from the perspective of an individual living within a society but not as a part of that society. (That is, the individual's concerns are given primary

importance; e.g. freedom of speech, which may well lead to negative social outcomes in certain instances of its use but which is thought to be of sufficient importance for the individual that it is guaranteed. Framers of such laws naturally see a society in which each individual has free speech as being a better society, but their focus in crafting such a decree is on how an individual will live, not on how the society as a whole (in a transcendental sense) will live.) Focused fundamentally on the individual, and her relationship with the greater society only in so far as she needs shielding from it and liberties to operate as she pleases in defined ways within it, this account takes form through the liberal policies and governance it enacts. Thus it is that Dworkin takes the essence of liberalism to be equal consideration of and respect for each individual, whose civil liberties are needed to guard them against the preferences of others concerning how they should act (Gaus 2000). Similarly, Narveson takes his 'liberty rule' to be of paramount importance, by which one may do whatever one pleases granting that such doesn't harm another or interfere with another's affairs. This attention to a person's actions and how the actions of others may affect said person fundamentally rests upon the extension of ownership between that owning and that owned; this is the self, understood as one's will or one's mind, owning one's body and behavior in the way a material object is owned (Narveson 1998). Responsibility for what is done and the resultant consequences therefore rests with each free-floating mind and the ramifications of any one person's acts extend only out to one degree of separation from their source. This is an atomistic self, one that is the 'owner' of the actions meant to be protected from incursion upon or held responsible for the effects of, depending on the situation, and since bodies are owned by minds and are the means by which actions are executed, the chain of ownership-responsibility is clear: action  $\Rightarrow$  physical body that committed action  $\Rightarrow$  mind that directed physical body. The self, to put it crudely, is that point where the buck stops.

A primary problem with this line of reasoning is that we still don't know what the self is. Is it an 'individual' – a collection of organic materials animated by a directing mind? Is it that directing mind itself? What exactly *is* a mind? Philosophers and neurologists have long discussed what a mind, as opposed to a brain, can be said to be and have still not arrived at a satisfactory definition, though it can be hoped that research into consciousness may one day settle the matter. Narveson, however, only adds to our confusion in his argument on body ownership, claiming that 'Everyone is "boss" over his own mind' (ibid., note 35, 18); if I'm myself and the self is my mind, and I'm boss over my own mind, then all we've done is to come back to our original query of what the self is. That there is nothing that can be pinned down, but there's still something there, is

what makes this the soft anti-realist position rather than its hard variant (to be discussed below). A number of ideas have been proposed to fill in the gap here, notably Dennett's and similar narrative accounts of the self as the central character (or 'center of narrative gravity') in descriptions of what happened to a given brain-body aggregate (Dennett 1991). This is still a something foisted onto a nothing, but does take into account the brain's parallel and distributed structure; we have no 'central command center' but instead have a number of systems that are specialized and locally processing, bound together in a network out of which the mind arises as an emergent property. Although we think that our consciousness is unified, it is in fact a 'constellation' of specialized consciousness systems whose products are thought to be integrated and interpreted by a cognitive module that evolved for that purpose (Gazzaniga 2011). (It may be tempting here to label the interpreting module as being what the self really is, but even if that module is definitively proven and its location – if it has one – is discovered in our neural mass, its function would remain descriptive and not generative.) Nevertheless, as Kristjánsson points out, soft anti-realist positions, be they of the narrative or another variety, cannot differentiate between self-knowledge and self-deception (Kristjánsson 2010): I may think that something happened to me only to later be told that that actually happened to my brother and I was misremembering what he had told me; before being informed of my error I was convinced that what happened to my brain-body was the contents of the story in question, and its central character was me. After learning my mistake, what is the identity of the central character in the erroneous section of my narrative? While I held the false memory were my self and my brother's self somehow the same, had his self entered my narrative? And where was my true self during the lost time in which my memory has deceived me into narrating falsely? Who was the 'me' of the events that really occurred? Without a self that can be tracked in any real sense we are left to wonder. Finally, this account, with its viewpoint of sealed off and atomized individuals, egregiously fails to notice that none of us exist in a vacuum and that each self is highly contextual, based not only on current situation but also local culture, historical time period, geographic location, and a host of other details (Ravven 2013). The self cannot be defined in the absence of such because the self will never exist in the absence of such, nor will the self ever be fully free from outside influences that affect its behavior and decision-making, a point we will return to. For now, we move our attention to the other anti-realist account, that of the hard version.

*B. The hard anti-realist position,  
or The reductionist nonself*

The hard anti-realist position endeavors to rectify one of the central errors of its soft cousin: that of attempting to be a something with nothing underneath defended by those ‘who confuse the masks with the face’, as Kristjánsson memorably puts it (Kristjánsson 2010, 44). This account is laid out in great detail by Parfit in his *Reasons and Persons* (Parfit 1984), there called the reductionist view. Parfit uses a series of imagined scenarios of varying degrees of likelihood (such as a false memory about a day in Venice or a teleportation to Mars gone wrong) to arrive at and then expand on the conclusions he uses to define the hard anti-realist viewpoint: 1) we do not exist separately from brain/body, physical/mental events, 2) identity is not always determinate, 3) unity of consciousness and the unity of a whole life cannot be explained by claiming many ‘different experiences are had by the same person’ but must describe the relations between the experiences and the person’s brain and can be fully described without claiming the ‘experiences are had by a person’, and 4) personal identity doesn’t matter and what does matter is psychological connectedness and/or continuity with my cause for such (ibid., 216-217).<sup>1</sup> This final conclusion – ‘what does matter’ – is labeled relation R, which Parfit defines succinctly as: ‘psychological connectedness and/or psychological continuity, with the right kind of cause’, adding ‘in an account of what matters, the right kind of cause could be any cause’ (ibid., 262). It may be helpful at this point to introduce an example of what Parfit means, and so we will look at one that he himself uses, namely, teleportation gone wrong. Imagine that you have been sent to Mars by your employer on an assignment, but due to the limitations of space travel the fastest way to get there is not to be physically transported but instead to use a teleportation device,<sup>2</sup> a machine that doesn’t actually move anything but rather reads it and reproduces it elsewhere. The way the teleporter works is that it scans the current state of every cell in your body and then fully reconstructs an exact replica of your body at its sister location on Mars. (Parfit considers both cases of the machine destroying your body on Earth in the process of replication and also that of your Earth body not being destroyed; his discussion of the two body scenario is interesting but beyond our current scope.) In such a case all that you are left with in your brand new body, but still exactly like your old body, is relation R, but this does

not mean that the self is relation R because, again, there is nothing there of substance post-teleportation that has remained. There is no self, and we deceive ourselves into thinking that there is only because we typically have both relation R and physical continuity; what Parfit means to demonstrate is that there are cases (albeit at present imaginary ones) that challenge these presumptions but still force us to admit that I am my Mars ‘me’ without there being anything of ‘me’ on Mars at all. Parfit finds this very liberating, proclaiming ‘On my view, what fundamentally matters, in our concern about our own future, is the holding of relation R, with any cause. This would be what matters even when it does not coincide with personal identity.’ (ibid., 289) Instances where relation R does not coincide with personal identity would be like that just discussed, a ‘branch-line case’ where your Earth body stops but your Mars body begins, or a ‘division case’, where your Mars body begins even while your Earth body goes on, making both bodies ‘you’; and since there is no self, both bodies *can* simultaneously be ‘you’. This potentially takes the sting out of death – it evidently does for Parfit, anyway – and allows us to view our own mortality as just one more blip in the long stretch of the natural world’s cycle. There is nothing that is essentially me, so when I die nothing will be lost. Indeed, by this view, we may ‘die’ any number of times during our lives as relation R is lost and taken up again (e.g. your early childhood years of which you have no recollection when shown photographs of the time, or even blacking out what happened after your staff party the night before). Parfit tells us, ‘If we are Reductionists, we regard the rough subdivisions within lives as, in certain ways, like the divisions between lives. We may therefore come to treat alike two kinds of distribution: within lives, and between lives.’ (ibid., 333-334) In its full denial of any self the hard anti-realist account may be counterintuitive yet it still appears to be clear and convincing.

It is not, however, without its problems. To begin with, there is the matter of personal identity and Parfit’s claim that relation R matters more than it. In considering what we think we are, that is, what our personal identities are, Parfit examines both physical and psychological criteria and declares both to actually be reductionist. The physical case is that a person is the same person if enough of the brain, and not the whole body, has continued (without branching) between the past and the present; the psychological case is that a person is the same person if there is overlapping psychological connectedness that forms a psychological continuity between the past and present, also without having branched. These both boil down to Parfit’s view, he writes, because: 1) ‘the fact of a person’s identity over time just consists in the holding of certain more particular facts’ and maybe also 2) ‘these facts can be

1 Parfit also allows that physical continuity and similarity may be important too.

2 Parfit actually calls it a ‘teletransportation’ device; I have shortened it for simplicity’s sake.

described without either presupposing the identity of this person, or explicitly claiming that the experiences in this person's life are had by this person, or even explicitly claiming that this person exists. These facts can be described in an *impersonal* way.' (ibid., 210, emphasis in the original) We must therefore reject both of these to be a non-hard anti-realist regarding the self, Parfit claims, and instead take personal identity as involving something beyond what the physical and psychological cases are propounding: a further fact of the soft anti-realist or realist kind. We can grant Parfit the point that the physical and psychological arguments reduce to his reductionist position without agreeing that relation R ('psychological connectedness and/or psychological continuity, with the right kind of cause') has more value than personal identity because what each of us takes our identity to be is singular and resonates with deep personal – and only personal, it cannot be transferred – meaning for us. This meaning is the 'something beyond'; this is the 'further fact' connecting the present person with their future version beyond how R is defined. Parfit tries to preempt this uniqueness objection by the following: He concedes that when R is held uniquely (U) in a one-one form (i.e. nontransferable (and nonbranching in teleportation or other imaginary cases)), then personal identity (PI) is equal to relation R plus uniqueness, thus:  $PI=R+U$ . Parfit then reduces U by stating that if I am R related to a person 'the presence or absence of U makes no difference to the intrinsic nature of my relation to this person', and hence adds very little to R, although it 'can be plausibly claimed to make a small difference'. But this 'small difference', Parfit then states, 'would be much less than the intrinsic value of R. The value of PI is much less than the value that R would have in the absence of PI, when U fails to hold.' (ibid., 263) This is technical and slippery writing, but if looked at closely it reveals a large hole in the logic of the argument that is clear when we substitute in some whole numbers. We will follow Parfit's estimation of the unimportance of U and make that a 1, and we will also follow Parfit's estimation of the importance and high intrinsic value of R and make that a 4. There is no way that PI's value would be 'much less than the value that R would have in the absence of PI, when U fails to hold' because all of these elements are related and R in the absence of PI is the same as R on its own, i.e. R before U has been *added to it to make PI*. In our numeric substitution we have  $U=1$ ,  $R=4$  and therefore  $PI=5$  ( $PI=R+U$ ). Take U away from that so now we have R on its own and we still have  $R=4$ , but PI is not less than R and nor is it equal to R, it is nothing at all because we need to add U to R to arrive at PI. (PI does not equal 4 here since if U had *no* value then  $PI=R+0$ , which is the same as  $PI=R$  and we have already established that  $PI=R+U$ .) Even if R is not held uniquely, say, in the branching case where

Earth you and Mars you both exist, each version of you would still have a personal identity that meant a great deal to each and that would diverge from the other-planetary double's personal identity with each passing second as their lives moved on and experiences added up.<sup>3</sup> We now have a strong claim contrary to Parfit's assertions that relation R matters more than personal identity and that personal identity does not matter at all (recall his initial conclusion 4 above: 'personal identity doesn't matter and what does matter is psychological connectedness and/or continuity with my cause for such') that we can add to the more commonsense objections to the hard anti-realist position of which Parfit's account is representative. Foremost amongst these, as the preceding has endeavored to show, is that by this view things simply don't work. Kristjánsson writes that hard anti-realists 'take pot shots at the notion of truth as correspondence with reality. There is a basic difficulty with rejecting this notion: Almost all human actions, communications, interactions and investigations seem to presuppose its truth.' (Kristjánsson 2010, 38) (That Parfit's arguments rely on scenarios that themselves have no correspondence with reality may be instructive in this regard, but their usefulness as analytic tools could belie that.) Furthermore, without a sense of persisting internal sameness that lasts over time, moral responsibility becomes increasingly difficult to pin down (ibid.). If we can indeed treat subdivisions within lives as being like different lives in some ways then we may find ourselves embarking on a path that results in entirely undesired reforms to our justice systems (e.g. is Frank in his current life responsible for the crime the entity we used to call Frank – who looks just like our Frank now but who experienced such a subdivision – committed prior to the subdivision? What if he is found not to be? Other undesired judicial changes stemming from different reasons are also a possible outcome of the contextualized soft realist account discussed below). On the whole, the hard anti-realist viewpoint 'makes travesty of everyday moral experience' (ibid., 46), and it is also unclear how the ubiquitous situational and contextual pressures that we all face would affect us on this account, or indeed, if they would be thought to affect us at all. Finally, as we noted above, emergent properties can and do exist physiologically: the sense of a unified conscious mind is

3 Relation R would also diverge as time went on, of course, eventually leaving each double with potentially no connection to their original singular being (as in the case of the nonrecollection of childhood events). The personhood of each is in question here, and that issue is debatable (note though that in both the physical and psychological criteria there is a 'no branching' rule), but the point that I am trying to make is not about personhood, only that relation R does not matter more than personal identity.

a primary example of such. Could the self not therefore also be an emergent property of our natural biological functioning, even if it cannot be pinned down in a hard sense? The next two subsections will consider the self from that and related perspectives.

*C. The soft realist position,  
or The Humean emotional self*

What will perhaps strike many readers as being most remarkable about the soft realist account of the self is that it is realist – that it claims that there is in fact a substantive nature to selfhood without making the further Cartesian ego or soul claims that we tend to associate with the realist stance. While admitting that there is no consensus on the definitive way to interpret Book II of Hume's *A Treatise of Human Nature*, Kristjánsson suggests that Hume 'seems to be arguing that, whereas the self as a succession of related ideas and impressions cannot be a direct object for the understanding, the self of whose moral actions each of us is intimately conscious can be a direct object for our emotions', making the self's realism, its actuality, consisting of emotional activity and generating a self-concept that is dependent on reinforcement from others via social interactions and societal rules and conventions regarding emotions (ibid., 47–48). The self here is each person's moral being, the day-to-day psychological unit of reference we go by, akin to 'the voter' or 'the citizen' or 'the taxpayer'; it is seeing oneself from an affective and morally related point of view (ibid.). Kristjánsson summarizes his soft realist self as being composed of three sets of self-related emotions: 1) self-constituting emotions: those that define us, our 'core commitments, traits, aspirations or ideals', 2) self-comparative emotions: those that take the self as 'an indirect object' or 'a reference point' for 'comparison with a baseline of expectations', and 3) self-conscious emotions: those that are in the self they are about, that take the self as 'their direct attentional and intentional object' (ibid., 75–77). The self here is that which we carry around with us and that more or less matches what we mean when we speak of ourselves, it is the culmination of the creature performing the actions, thinking the thoughts, and having the feelings that we internally associate with those actions, thoughts, and feelings, and that others associate with them too. This is the self that we're used to in the hard realist sense but without any notion of an element that is eternally existing or inhabiting from the outside. There is no mental 'pure ego' here that takes up residence in the physical body, rather the physical body, along with our unique sense of personal identity, is all that the self is and part of that body's normal functioning is to have the emotion-based ongoing characteristics that make up Kristjánsson's first emotive set above and the self-referencing features that

make up his second and third sets. This view of the self, like the narrative variety of the soft anti-realist position, matches with the multivariate and highly specialized modular functioning of our physical brains (Gazzaniga 2011), but unlike the narrative account Kristjánsson's soft realist self posits a substantive object (the day-to-day psychological unit) that can differentiate between self-knowledge and self-deception, even if it cannot be said to exist physically. Our self-concept as a moral being, moreover, means that it can correspond to reality or fail to correspond, it can be judged objectively through the lens of its quality as other-dependent in the social realm. Our daily interactions will either reinforce the way in which we see ourselves as being or they will show us where we have been in error.

Of the accounts of the self examined so far this one is the strongest in both the theoretical and neurologically accurate (as regards functioning) senses, yet it still leaves something to be desired and this has to do with how Kristjánsson handles agency and decision-making in his soft realist self. In a discussion of the 'gappiness problem' taken from psychological research (Blasi 1980), where it has been found that moral reasoning either fails to motivate moral action or does so only slightly, Kristjánsson refers to his 'unified moral self of rationally grounded emotion' as a means of repairing the disconnect – or rather, showing that there isn't a disconnect and that the root of the problem of failed moral action lies elsewhere (Kristjánsson 2010, 97). Kristjánsson takes the soft realist self as showing that there is no difference between the 'moral-self' (having moral concerns as part of one's identity (Blasi 1980)<sup>4</sup>) and having moral emotions, and thus motivating moral action, because those emotions are the foundations of the self. Moral emotion can join together moral cognition and moral action through the training, in an Aristotelean sense, of a (soft realist) self to be a moral self because there is a difference between episodic emotions and dispositional emotions, with the moral (soft realist) self having its grounding in the latter. Emotional reactions show the internalization and integration of 'a certain emotional disposition into his or her moral self.' (Kristjánsson 2010, 94) In a sense, I think that Kristjánsson's account here is accurate, yet it fails to note the evolutionary grounding that our emotions have and gives rationality a more central role than it appears capable of taking, stating that a reflective decision is required to make moral concerns part of one's self-identity and that, again, the baseline of a moral self is 'rationally grounded emotion' (ibid.). There is a large amount of empirical research that demonstrates that most of our decisions are in fact made unconsciously, based on

4 This was Blasi's suggestion for how to solve the gappiness problem.

intuitions that have been honed by natural pressures over millennia, and that only after a decision has been made and acted upon (or thought or feeling generated) does our rationality step in and provide an internal reason for the action/thought/feeling (Haidt 2001, 2012).<sup>5</sup> The intuitive rules that drive our decisions seem to be both *prima facie* and reflexive, biologically useful for their efficiency even if not always correct (Osman & Stavy 2006). Gazzaniga summarizes the brain's functioning in this way with 'Many moral intuitions are rapid automatic judgments of behavior associated with strong feelings of rightness or appropriateness...not usually arrived at by a deliberate conscious evaluative process that has been influenced by reason' (Gazzaniga 2011, 166). Some of our decisions can nevertheless be swayed by reason, particularly when coming from others and thereby having a social affect on us (Haidt 2001), and our intuitive evaluations can change during our lives, but for the most part by the time we have gotten to the point where we are rationally deciding what to do, we are not – contra Hare – coolly considering our now-for-now and now-for-then preferences and then weighing them against logic and the facts (Hare 1981), we are instead attempting to explain to ourselves the decisions that our unconscious minds have already come to and put into play. Kristjánsson's account is not the fully rational one that Hare's is; it is, after all, based on emotion and seeks to promote individuals who will thoughtfully decide what kind of emotional dispositions they seek to have and will then set about inculcating them through the chosen or designated method. In many ways this is quite similar to Haidt's model, and it is in Kristjánsson's rejection of that model that we can see the Achilles' heel in his own. Kristjánsson dismisses Haidt's and similar systems, as well as the opposing rationality-centered systems, as being two-tiered (emotion on one level and reason on another), where 'emotion is distinct from – if complementary to – reason' (Kristjánsson 2010, 98) and thus as not having placed one within the other, one as the other: his 'rationally grounded emotion', as above. Yet if the research revealing the timing of our decision-making (intuitions first, reasoning later) is correct, and at this point it certainly seems to be, then perhaps Kristjánsson's order should be reversed, giving us a soft realist self of emotionally grounded reason. That would, however, still leave Kristjánsson's concern with the emotional training of a moral self intact, a point with important ethical ramifications.

#### *D. The contextualized soft realist position, or The embedded and determined self*

The final self account that we will examine is similar to

Kristjánsson's in being a soft realist position, but is more inclusive of scientific research into cognitive functioning in its theoretical framework and far more radical in its conclusions. This is Ravven's contextualized soft realist position, applying a label that neither Ravven nor Kristjánsson has given to her work and one that I hope will not be misleading. This account also posits a substantive self that is grounded in emotion but stretches that foundation to layers far below where Kristjánsson has them. Ravven cites the work of Panksepp (a neuroscientist) which argues that the self emerged evolutionarily as an affective system to facilitate survival, that it stems from very early sections of the mammalian brain, and that it is something that is shared across all species of mammal (Panksepp 2011, in Ravven 2013). As such, the self here is our (and nonhuman mammals') 'point of view of survival' and 'first emerges in the precognitive ability of most organisms to operate from an ego-centric point of view.' (Asma & Greif 2012, in Ravven 2013) Although this is an emotionally-based self, it is not a Humean reflective self composed of the three sets of self-related emotions that Kristjánsson detailed; rather, it is more of a primal urgency, a voice from deep within that guides and directs. It is here that this view's highly contextualized nature can be most clearly seen, for not only is Ravven's version of the soft realist self a substantive one, it is a highly situationist and mostly deterministic one. Ravven stresses the embedded nature of the composition of our sense of self, stating that it is constructed by its current relation to another; that is, that we all have multiple selves each of which corresponds to a significant relationship and is partially formed by that very relationship and its object (other) through the carrying over of the sense of self involved. Despite this position's status as a soft realist view, none of us have a singular 'me', but instead 'the feeling of self is a mental capacity that can be projected inward or even outward onto the world...we make parts of the world feel like self, and we fill our feeling of self with our engagements in the world.' (Ravven 2013, 372) This notion of a malleable, distributed self that twists and turns with the forces around it also leads Ravven to write that our actions and fate 'are determined... by who our parents were, what world and situation we were born into, and who we became as a result of our early experience, our genetic inheritance, and on and on' (ibid., 348). Accordingly, the agency we assign ourselves mistakenly infers causal ownership of our actions when in fact there lays behind each act a multitudinous number of causes and conditions, that we ignore contributing factors and falsely imagine our behavior to spring from an unbiased free will (ibid.). Ravven does nonetheless try to assign moral responsibility by following Lear's lead in his analysis of Oedipus: that, despite all that may have been ordained by outside forces, what matters in the end

<sup>5</sup> Both sources summarize the evidence and present Haidt's Social Intuitionist model in detail.

is that it was done by us. Ravven writes that ‘This given “me” is that by which I am constituted. It is the “me” I find, and I resign myself to accepting it. In so doing, Lear says, I become transformed from being passively acted upon into a morally responsible agent.’ (ibid., 348-349; Lear 1998)

There is much that is good in the socially minded and contextualized aspect that informs this position of a soft realist self, taking into account as it does the lessons we have learned of the overwhelming influence of group and setting from such famous psychological experiments as Zimbardo’s 1971 Stanford Prison Experiment and the electric shock experiments studying the tendency for obedience to authority figures by Milgram, both of which Ravven refers to on numerous occasions. Moreover, factors such as our upbringing, historical time and place, socioeconomic background, and the capriciousness of the genetic lottery all certainly play large parts in our lives. However, to assign full determinism is a step that many will see as going much too far, and Ravven does more or less do so, if not in so many words. Her version of agency and moral responsibility, for instance, amounts to little more than an admission along the lines of ‘This thing that I did was generated by uncountable and interrelated background causes over which I had no control, culminating in the performance of the act by my physical body, but I’ll accept the consequences of the action anyway.’ Very few of us would consent to taking on responsibility in this way if we held such a view, particularly in legal contexts, and if this account is true then our legal systems themselves would need an overhaul of a proportion that is difficult to even imagine. Gazzaniga, in his consideration of social influences and the brain’s highly programmed and often unreachable workings that occur automatically and unconsciously, still concludes that ‘ultimately responsibility is a contract between two people rather than a property of a brain, and determinism has no meaning in this context... Brakes can be put on unconscious intentions’; conclusively ‘we have to look at the whole picture, a brain in the midst of and interacting with other brains, not just one brain in isolation.’ (Gazzaniga 2011, 215) Despite the long arm of our genetic inheritance, and all of the factors outside our grasp, we are creatures with agency who do have at least some degree of control over what we do, although that degree may vary widely from person to person and even from time to time within one person’s life. Ravven replies to Gazzaniga’s position on this by citing his lack of consideration of neuroplasticity – that our brains’ neocortical pathways are rewired by experience – and that because of that neural characteristic the influence of culture, meaning, and language are not voluntary but are flexible; we cannot change our patterns of thought through will but ‘only by training and re-training’ (Ravven 2013, note 96, 470-471). This appears to leave

the door of choice open a crack, at least as far as our situational interpretations go, which would surely play a large part in subsequent behavior. But Ravven on the whole appears to be uncomfortable with that stance, preferring instead the more potent near-determinism that marks her discussions of the changing self, the contexts in which we move, and the primacy of the ego-centeredness that is an ancient evolutionary heritage fueling our affective systems which in turn drive our behavior. This is Haidt’s social intuitionism, which Ravven also references, as well as Greene’s work (Haidt 2001 & 2012, Greene 2013),<sup>6</sup> with little or no hope for a self-reflective feedback loop – a position that I would be very surprised to find either psychologist holding. In her book’s concluding section, ‘A Final Word on Moral Responsibility’, Ravven even goes so far as to state that ‘If free will is relinquished, we come to recognize that what must be *must indeed be*, and that what must have been *could not have been otherwise*.’ (Ravven 2013, 419; emphases in the original) This is a very comforting thought to all of us who have regrets, but if true then it would take us into a realm of unalterable fate that not a few would find abhorrent and it would also wreak havoc on the meaning and purpose in life that many people have discovered for themselves. There is a final problem with this account as well, and that is that Ravven appears to take what are still controversial findings as conclusive. For example, she cites Damásio’s position that the self defines itself by its relations to the environment and its relationships with others (referencing Damásio 1994 and 1999) as being a fundamental and crucial point, yet she also states that the data is at present inconclusive. Panksepp’s work is also controversial, as is the extent to which Ravven embraces neuroplasticity. When drawing our conclusions from the research being done into the brain we must remember that the field of neuroscience, along with its revelations about cognitive functioning, is a rapidly changing one and this is something that Ravven does not seem to keep in mind. Although her ideas are exciting and, if they withstand the test of time, would potentially mean a radical reorientation of human life, it seems to me that Ravven’s contextualized soft realist self would be far stronger and more applicable at present if it were tempered down by a large margin. Nevertheless, in what follows we will take much from this account, as well as from Kristjánsson’s soft realist self, in our offering of a self-concept that seeks to accurately reflect the conclusions about our natural functioning that are widely accepted, and that may allow us to approach one another in a more ethically affirmative manner.

6 Both of these researchers present the brain’s functioning as being prewired and specialized by area – the modular view that is at present generally accepted but which Ravven rejects in favor of a more plastic view.

### 3. The Limited Choice Soft Realist Self

There does seem to be good reason for accepting the commonsense view that there is substance to the self although that substance is almost certainly not in a physical or definitive form; even the soft anti-realist account is based on this assumption – with its something that is supported by nothing – despite being an anti-realist position (though the hard anti-realist position is an exception here). Ravven's account of an emotional self made up only or primarily by a biological drive to survival and that exists as a multiplicity which transforms depending on current relational statuses seems excessively shallow, however, and it may be noted that its foundational survival drive can be included in Kristjánsson's fuller three sets of self-related emotions, fitting into the third category of self-conscious emotions. It is Kristjánsson's account generally, in fact, that seems to be the most robust of the four that were examined, and its deficiencies can be made up for by supplementing the position rather than by having to start over again from scratch. To his affective and self-reflective account we may add the growing psychological evidence in favor of an intuition first, reasoning second model of decision-making and behavior initiation which is augmented by Ravven's emphasis on background causal factors and social pressures playing an important role in shaping the parameters that influence our intuitions. As Rorty has put it, 'the central flaw in much traditional moral philosophy has been the myth of the self as nonrelational, as capable of existing independently of any concern for others' (Rorty 1999, 77), and to this relational aspect may be added the historical, geographical, and epochal embedded nature of the self that Ravven highlights. The place in which we find ourselves, in which our selves exist, is one that is marked by many outside pressures and unconscious leanings, our selves are interdependent with each other and with the environments which we inhabit. Our selves may not *define* themselves this way, as Ravven cites Damásio as arguing, but it nevertheless seems reasonable to hold the lesser position that how others see us does play into how we come to see ourselves, as per Kristjánsson's Humean model. Moreover, these very multiple layers of mutually dependent and interlocking social fabrics and influencing factors that form our world point to a limiting of potential actions and options, and the more so when we consider that in most cases our unconscious minds will be the deciding element, with their own internal influences of current affective aspects, personal values held, life experiences, and the like. We are not fully free to do absolutely anything at any time because what we will be able to do, and more to the point, what we will consider ourselves able to do, will be limited by factors

over which we have no control. The fact that everyone else is in this same situation should give us pause when starting to judge others as we recognize the limiting external and internal restrictions that we all face. This is not a universalist ethics, but it does acknowledge that the broad generative processes that affect our choices as individuals are shared across our species, differing by regional, cultural, and linguistic inputs but still affecting us in similar ways. That our choices are limited does not, however, mean that they do not exist, and it is because of their existence that we retain agency and moral responsibility. As Gazzaniga points out, criminals do not usually commit crimes in front of police officers (Gazzaniga 2011), and everyday experience confirms both that we have self control and are able to exercise it to varying degrees, unless one is willing to accept the full determinist position and consider our perceived self control to be an illusion; some problems with hard determinism have been mentioned above and are discussed at length elsewhere. Our intuitions and social pressures may be pushing us in a certain direction but we still have the ability to choose within those parameters that we confront, and more importantly, we also have the option of partially choosing how we will intuitively react by working to foster positive immediate unconscious reactions and minimize negative ones through reflection and repetitive training.

Although some arguments against a soft realist view of the self have already been discussed above in our consideration of the representative anti-realist accounts, perhaps the most direct objection that could be made would be that it is simply one more theorized account among many, with no hard evidence that would cause us to consider it more accurate than its rivals. While Kristjánsson's three sets of self-related emotions that compose the soft realist self may be difficult to ascertain empirically rather than experientially, our shifted foundation to that of an emotionally grounded reason – instead of the other way around – that we gave to his self-concept in light of the research on how our decision-making tends to happen (intuitive reactions primary, rational thinking secondary) does have a sound basis in the experimental sciences. Greene notes that 'From a neural and evolutionary perspective, our reasoning systems are not independent logic machines. They are outgrowths of more primitive mammalian systems for selecting rewarding behaviors – cognitive prostheses for enterprising mammals. In other words, Hume seems to have gotten it right.' (Greene 2013, note to 137 text, 368) Since our ability to reason, and likewise all that we take to be quintessentially human about that, has at its core an affective network of action promotion, similarly grounding our internally held view of self in the emotions would appear to be far more valid than any of the alternatives, and unless a hard realist self were to

somehow be discovered this seems to be our best bet.

What ethical applications may such a view of the self entail? To begin with, by focusing on the embedded and interconnected nature that all of our selves have it would be easier to see in others commonalities with oneself. This would help us put ourselves in others' places when interacting with them through the recognition that they too are affected by similar social and related pressures and that their actions and outlooks towards us arise from a limited range of options. It would also inform the way we regard others with whom we have little or no direct contact by placing them within their respective wider spheres vis-à-vis our own, and understanding that there are many generational pressures culminating in current practices – something that could be particularly advantageous if applied to current debates on multiculturalism and the struggle for acknowledgement that minority groups often undergo (Hanssen 2000). A real-world example may help to illustrate the ideas currently under consideration, and to make the case a little more interesting we will use an incident from my own life here in Tokyo which includes intercultural elements that might otherwise not be present but in light of the preceding multicultural comment and our increasingly globalized world seem appropriate. As for many Tokyoites, taking the train occupies a central part of my life and dealing with the crowds and the stress of the daily commute presents many opportunities for conflict. Not long ago I was taking the stairs up from the line I had been riding on to transfer to another line and continue my morning journey when an incident that relates to the present discussion took place. Typically one is expected to walk on the left side of the staircase when ascending, but there are many stations where signs have been posted directing people to instead go up on the right side and down on the left (as you face the stairs from the bottom); the station where I transfer is one such place. On that day as I started to climb I noticed that a man was coming down on the same side that I would go up on, but as the signs said that where I was was designated for ascending I assumed he would move over and continued going up, keeping my eyes on the steps immediately in front of me to avoid tripping. Suddenly I saw his feet on the next step and looked up to see him standing there glaring at me. He had not moved over at all and, judging from the look on his face, was very irritated that I was on the side of the stairs that I was. For a split second I considered pointing to the very visible signs stating that this was the ascending side both behind me on the lower floor's ceiling where the train line information was posted (visible as you descend onto the lower floor) and on the stairs themselves while saying something polite but direct, but then thought better of it and moved around him, glaring in return as I did so.

Now, I admit that I could have acted much better in

this situation, but I also could have acted much worse had I not recently been thinking about the affective forces and background issues in people's lives that influence behavior; to see how all of this fits together allow me to unpack this brief episode using the foregoing analysis. To begin with are Kristjánsson's three sets of (my) self-related emotions and how the everyday contributes to continually shaping the self: 1) my self-constituting emotions (defining commitments, traits, ideals) include a strong sense of justice, that what is right ought to be favored and that this includes the proper following of established social rules designed to make everyone's life safer and better (e.g. traffic laws, deference conventions, etc.), this led me to believe that my following the directing signs on the stairs should be reciprocated by others; 2) my self-comparative emotions (referencing the self indirectly against baseline expectations) involve an awareness of my status as an outsider here in Japan and the pressure I feel not to stand out any more than I already naturally do; viewed against my own expectations regarding how I think I should behave I 'rated' myself in this instance fairly poorly; taking us to 3) my self-conscious emotions (in the self they are about, attentional and intentionally direct) informed me that I could have and should have done better, I realized too that I was tired and irritable that day and that those feelings played into it; I also realized that the other man may have been tired and irritable as well, contributing to how he behaved. This last point also demonstrates how the intuitive and affective backgrounds that we have feed into and connect with our self-related emotions and help unconsciously form the boundaries of the spectrum of choice that we consider ourselves to have in those instances when we stop and apply our conscious minds to the situations we find ourselves in (rather than just operating unthinkingly from the intuitive level). These boundaries of choice are in addition restricted by any number of contextual elements. With my own North American cultural background my intuitions (functioning automatically and unconsciously) told me that I had no need to – and indeed should not – move over when I first glimpsed the man descending on the wrong side: people who are in the right do not yield to those who are mistaken. Surely he would see the signs and move to where he should be, it would be best to carry on where I was. Although I cannot speak for the other man involved his intuitions may have been giving him pre-packaged decisions (as it were) indicating that one must always walk on the left side, or that as a foreigner I ranked below him on the social hierarchy scale and therefore should submit to him, or that as someone younger I should allow him his preference (I am often told that I look younger than I am, something that in many places would be advantageous but works against one in a society where age is afforded status).

Each of us could have overridden our intuitions but we failed to, and I could have met him with a smile and a laugh as I moved around him instead of a returned glare but I did not; this is a point I should address as I seek to equip myself with improved intuitive reactions, another lesson that can be taken from the psychological research that underpins the self-concept being outlined here.

The above is just a tiny example but it does give us a glimpse into how the process of applying this view of the self ethically could work, and how our awareness of these issues and our monitoring of our self-related emotions can make us behave better towards other selves and our own selves. The self as understood on this view is something that is an ongoing and real part of us, but it is also constantly being shifted, reinterpreted, and reconstructed as we go about our lives, reflect on ourselves, and receive feedback from others. Moreover, this is happening for every self in our vastly interconnected human milieus, each interaction adding to the picture of the self for that self. If we can therefore foster a deeper awareness of and respect for the interdependent links that we share with the other selves both within and without the public arenas in which we operate then we may be able to approach an understanding of what we are that includes many more of those others than current atomistic accounts permit. Rorty writes that 'Moral development in the individual, and moral progress in the human species as a whole, is a matter of re-marking human selves so as to enlarge the variety of the relationships which constitute those selves.' (Rorty 1999, 79) The limited choice soft realist self-concept would move us in that direction, while also yielding a view of the self that better fits both our intuitions of what we are and the research into our cognitive constitution and behavioral output.

### Acknowledgements

A version of this paper was presented at the 'Self and (its) Realization(s)' conference hosted by Hokkaido University in May, 2015. I would like to thank the participants of that conference for their many thoughtful and useful suggestions. I would also like to thank the editors and anonymous peer reviewers of the *Journal of Applied Ethics and Philosophy* for their help with an earlier version of this paper. I hope that my efforts at improvement reflect the efforts they kindly extended on my behalf.

### References

- Asma, Stephen T. and Greif, Thomas (2012), 'Affective Neuroscience and the Philosophy of Self', *Journal of Consciousness Studies*, 19 (3-4): 6-48.
- Blasi, Augusto (1980), 'Bridging moral cognition and moral action: A critical review of the literature', *Psychological Bulletin*, 88 (1): 1-45.
- Damásio, António (1994), *Descartes' Error: Emotion, Reason, and the Human Brain*, New York: Putnam.
- Damásio, António (1999), *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*, New York: Harcourt Brace.
- Dennett, Daniel C. (1991), *Consciousness Explained*, New York: Little, Brown and Co.
- Gaus, Gerald F. (2000), *Political Concepts and Political Theories*, Oxford: Westview Press.
- Gazzaniga, Michael S. (2011), *Who's In Charge?: Free Will and the Science of the Brain*, New York: Ecco Press.
- Greene, Joshua (2013), *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*, New York: Penguin Press.
- Haidt, Jonathan (2001), 'The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment', *Psychological Review*, 108 (4): 814-834.
- Haidt, Jonathan (2012), *The Righteous Mind: Why Good People Are Divided by Politics and Religion*, New York: Pantheon Books.
- Hanssen, Beatrice (2000), 'Chapter 6: Ethics of the Other', in *Critique of Violence: Between Poststructuralism and Critical Theory*, London: Routledge, 186-209.
- Hare, R. M. (1981), *Moral Thinking: Its Levels, Method, and Point*, Oxford: Oxford University Press.
- Kristjánsson, Kristján (2010), *The Self and Its Emotions*, Cambridge: Cambridge University Press.
- Lear, Jonathan (1998), *Love and Its Place in Nature: A Philosophical Interpretation of Freudian Psychoanalysis*, New Haven, Connecticut: Yale University Press.
- Narveson, Jan (1998), 'Libertarianism vs. Marxism: Reflections on G. A. Cohen's "Self-Ownership, Freedom and Equality"', *The Journal of Ethics*, 2 (1): 1-26.
- Osman, Magda and Stavy, Ruth (2006), 'Development of intuitive rules: Evaluating the application of the dual-system framework to understanding children's intuitive reasoning', *Psychonomic Bulletin & Review*, 13 (6): 935-953.
- Panksepp, Jaak (2011), 'The Basic Emotional Circuits of Mammalian Brains: Do Animals Have Affective Lives?', *Neuroscience and Biobehavioral Reviews*, 35 (9): doi10.1016/j.neubiorev.2011.08.003
- Parfit, Derek (1984), *Reasons and Persons*, Oxford: Oxford University Press.
- Ravven, Heidi M. (2013), *The Self Beyond Itself: An Alternative History of Ethics, the New Brain Sciences, and the Myth of Free Will*, New York: The New Press.
- Rorty, Richard (1999), 'Chapter 4: Ethics Without Principles', in *Philosophy and Social Hope*, London: Penguin Books, 72-90.