**Inverse enkrasia and the real self**

Fernando Rudy-Hiller

This is a pre-peer reviewed version of the article published in *Thought* https://doi.org/10.1002/tht3.465

**Abstract**

Anti-reflectivist real self views claim that people are morally responsible for all and only those bits of conduct that express their true values and cares, regardless of whether they have endorsed them or not. A phenomenon that is widely cited in support of these views is inverse akrasia, i.e., cases in which a person is praiseworthy for having done the right thing for the right reasons despite her considered judgment that what she did was wrong. In this paper I show that anti-reflectivist real self views are problematic by focusing on the related but neglected phenomenon of inverse *en*krasia, which occurs when an agent commits wrongdoing by following a mistaken evaluative judgment that, unbeknownst to her, runs contrary to her true values and cares. Intuitively, inverse enkratics are blameworthy for their actions although the latter don’t express their real selves; therefore, anti-reflectivist real self views are false. I assess the implications of this result for the viability of the quality of will paradigm and conclude that the latter survives unscathed to the problems besetting real self views. The lesson is that defenders of the quality of will paradigm should stop talking about real selves altogether.

**1. Introduction**

According to a popular family of views, people are morally responsible for all and only those bits of conduct that reveal who they *really* are as moral agents—what their true values and cares are. This is so regardless of whether they have endorsed these values and cares (or even are aware of them) or not. Call this kind of views *anti-reflectivist real self* *views*.[[1]](#footnote-1) One sort of consideration that is often advanced in favor of them is that they are able to explain our intuitions in cases of *inverse akrasia*, that is, cases in which a person is praiseworthy for having done the right thing for the right reasons despite her considered judgment that what she did was wrong (Arpaly and Schroeder 1999).[[2]](#footnote-2) Since this person’s conduct speaks louder about who she really is as a moral agent—what she truly cares about—than her misguided judgments, praise is warranted despite the fact that she didn’t endorse the relevant motives.

In this paper I will show that anti-reflectivist real self views are problematic by focusing on a related phenomenon that has been neglected in the literature thus far: inverse *en*krasia. Inverse enkrasia occurs when an agent commits wrongdoing by following her (mistaken) considered judgment that her action was the right thing to do despite the fact that, unbeknownst to her, her deepest values and cares pulled her on the opposite direction. Intuitively, inverse enkratics are blameworthy for their conduct although the latter doesn’t express their real selves. Therefore, anti-reflectivist real self views must be false.

The upshot of my argument is neither that we must revert to older reflectivist real self views[[3]](#footnote-3) nor that we must reject the quality of will paradigm altogether[[4]](#footnote-4)—roughly, the view that praise and blame are justified simply by the moral regard (or lack thereof) people evince in their conduct. I do think that moral responsibility essentially depends on whether the agent’s conduct reveals something morally relevant about her. The key lesson of my argument, however, is that what is thereby revealed mustn’t necessarily be a *deep* fact about the agent. Expressing who we really are is surely an important aspect of moral life,[[5]](#footnote-5) but it isn’t nearly as important for moral responsibility specifically. I’ll suggest that we’d do better to stop talking about real selves altogether when dealing with the latter.

**2. Moral responsibility, quality of will, and real selves**

Moral responsibility is about judging and reacting to *people* for what they do. *Conduct* can be judged as right or wrong, but only people can merit praise or blame. But since it’s true that what triggers these evaluations and responses are always particular bits of conduct,[[6]](#footnote-6) there must be a morally relevant connection between the agent we evaluate and respond to and the bit of conduct in question. A widespread view, which I think is right, is that the relevant moral connection goes through the agent’s *quality of will*, understood as the regard (or lack thereof) for other people her conduct manifests.

 Once we have this proposal on the table, however, it’s tempting to take a further step and claim that what is thereby manifested must be a deep fact about the agent. Following Hume (1772/1978: 411), we might think that, given that actions are “temporary and perishing,” our moral appraisals and reactions must attach to something that is “durable and constant” in the person. It may thus seem that if the quality of will the agent’s conduct manifests is ephemeral or otherwise unrepresentative of who she really is, those appraisals and reactions would lack significance and depth. In this way we arrive at the idea that moral responsibility requires the expression in conduct of the agent’s real (or deep) self. Recent real self views are uniformly anti-reflectivist, since they develop an account of self-expression that eschews agentially demanding conditions like conscious reflection, endorsement, or identification with the relevant motives. Sripada’s (2016) account is the most detailed version of this sort of view, so in what follows I’ll focus on it.

 Sripada argues that the agent’s real self is constituted by her cares, which are a distinctive type of conative attitude functionally characterized in terms of a suite of motivational, commitment-related, evaluative, and affective dispositions. When an agent cares about something she has intrinsic desires related to the cared-for object; she is committed to sustaining that motivation should it come to fade over time; she is disposed to form favorable evaluative judgments about that object; and she is prone to experience a host of emotional responses triggered by the object’s up-and-down fortunes (1209-10). Whenever a person’s mental state exhibits this functional profile it’s ipso facto a care of hers, regardless of whether she identifies with or endorses it. Conversely, a person can misjudge her own cares: she may think she cares about something but if there is no associated mental state exhibiting the functional profile described above, then she doesn’t actually care for it (1211-2).

 Sripada then presents a “motivational support account” of expression, according to which the agent’s real self gets expressed if and only if one of the agent’s cares is among the motives that decisively (and non-deviantly) influence the occurrence of the conduct in question (1216). Finally, he makes explicit the relevance of the real self for attributions of moral responsibility by claiming, first, that “you are morally responsible for an action *only if* it expresses your deep self” (1212, italics added); and, second, that actions in which the agent’s real self fails to get expressed—for instance those produced by irresistible desires, irrational fears, or brainwashing—“are also cases in which, intuitively, the respective agents are not morally responsible for their actions” (1222).

 As I’ll show next, however, it’s false that failures of self-expression always defeat responsibility and, consequently, the necessity claim that Sripada (and every other real self theorist) defends is mistaken. Importantly, the agents in my examples do exhibit ill will and this is what makes them blameworthy. Thus, a further lesson will be that expressing a morally relevant quality of will doesn’t require the expression of the agent’s real self.

**3. Inverse enkrasia**

As I said at the outset, staple cases of the strength of anti-reflectivist real self views are cases of inverse akrasia. Unlike regular akrasia, inverse akrasia occurs when an agent performs an action that is actually right against her all-things-considered (mistaken) judgment that the right thing to do is something else. Huckleberry Finn has become the classic example of an inverse akratic. Huck helps Jim the runaway slave to escape despite believing he is acting wrongly in “stealing” from Jim’s “owner.” Moreover, Huck does it for the right reasons, namely because Jim is a human being just like him. Arpaly and Schroeder (1999: 163) claim that Huck is praiseworthy for his action precisely because it expresses his true self—“a good boy with his heart in the right place”—despite his objectionable moral beliefs. Other real self theorists concur with their assessment of the case (Doris 2015: 160-1; Sripada 2015).

 I’ll now present a class of counterexamples to anti-reflectivist real self views that is the mirror image of Huck’s case. These are cases in which agents, acting against their deepest values and cares—of which they, like akratic Huck, lack conscious awareness—are blameworthy for doing the wrong thing by following their considered judgment that their actions are right. Call this phenomenon *inverse enkrasia*. I’ll present three examples of it.

Imagine first an otherwise identical Huck who manages to overcome his deep resistance to turn Jim in and ends up doing so, thus aligning his conduct with his inauthentic moral judgment rather than with his authentic values and cares. Is this enkratic Huck blameworthy for his action? Intuitively he is, since he is a competent moral agent who intentionally does something that is actually wrong (just like the original Huck is praiseworthy for intentionally doing something that is actually right), and yet it isn’t clear how real self views could accommodate this intuition since, by hypothesis, his action doesn’t express his real self. For instance, the enkratic Huck’s action isn’t prompted by one of his cares as Sripada understands them (I’ll defend this point below) and, consequently, his turning Jim in doesn’t reveal who he really is as a moral agent.

Second, consider one of Sripada’s (2016: 1207) own examples of “a man raised under puritanical strictures [who] judge[s] that pleasures of the flesh are forbidden and to be avoided at all costs.” Sripada stipulates that “this evaluative judgment reflects the man’s acculturation ... It does not reflect the point of view of his self” because, although the man *thinks* he cares about sexual propriety, he is simply wrong about this: “There is no state in his mental economy directed at sexual propriety with the characteristic functional role properties of a care” (1212). Now suppose that this man does something wrong by following his inauthentic judgment that sex is sinful, for instance rudely rejecting the gentle romantic advances made by a charming acquaintance. Intuitively, this man wouldn’t be off the hook for his inconsiderate reaction simply because it didn’t express his real self. Again, however, real self views seem to be committed to deliver the opposite and implausible verdict that, because the puritanical man’s reaction isn’t self-expressive, he isn’t blameworthy for it.

Third and finally, consider a variant of another of Arpaly’s favorite examples involving “a young Ayn Rand fan who waxes ‘philosophical’ about selfishness as a virtue and the need to act selfishly but who herself is not selfish and acts unselfishly, and even altruistically, as a matter of course” (2015: 143). Suppose that on a given occasion in which this person notices the mismatch between her ethical views and her conduct, she musters sufficient willpower to act against what (unbeknownst to her) truly matters to her and scornfully refuses to volunteer for a local charity, taking this to be the Randian thing to do. Once more, the intuitive thing to say is that, regardless of how unrepresentative of her real self this action is, she is blameworthy for it. The real self theorist, however, is committed to rejecting this intuition on the grounds that the agent’s failure of self-expression blocks her responsibility altogether.

The counterexamples to real self views could be easily multiplied: just a find a case in which a competent moral agent does something wrong by following an all-things-considered judgment of hers that isn’t grounded on, and therefore fails to reflect, her true values and cares. Such an agent seems to be blameworthy despite the fact that her action doesn’t express her real self and, consequently, the two basic claims advanced by real self theorists are false: self-expression isn’t a necessary condition on moral responsibility and failures of self-expression don’t always defeat responsibility. At the same time, however, these cases aren’t counterexamples to the quality of will paradigm, since, by acting on their inauthentic evaluative judgments, these agents do evince something morally relevant about themselves, namely their taking certain considerations as a sufficient reason for acting in ways that fail to display proper regard for others (if only on a single occasion). To act in this way is to reveal an objectionable quality of will—however unrepresentative of who the person really is—and thus to become an appropriate object of blame.

**4. Objections and responses**

I turn now to consider several objections to my argument on the real self theorist’s behalf. First, it could be objected that the proviso I made when introducing the enkratic Huck example—“an otherwise identical Huck”—is unacceptable because, given the original Huck’s motivational structure, he would never have turned Jim in (and the same could be said about the akratic Randian). In response, notice that the inverse akratic’s praiseworthiness in the original examples isn’t premised on the assumption that her action emanates from a “volitional necessity” of hers, that is, that she is truly incapable of doing otherwise than she actually does. Rather, she is praiseworthy simply because, in the particular action at hand, she does the right thing for the right reasons (Arpaly 2015: 143). So inverse akratics needn’t be conceived as being motivationally unified. In fact, there are good reasons for thinking they aren’t, since they sincerely hold evaluative judgments that conflict with their true cares and values and which therefore (under the plausible assumption of judgment internalism) issue in motivations that oppose the latter. Thus, the possibility of inverse akratics acting enkratically on a particular occasion can’t be simply ruled out by appealing to the original description of the examples.

 A second objection concedes that inverse akratics possess conflicting motivations but argues that this fact shows that the actions of their enkratic counterparts do express their real selves after all. Anti-reflectivist real self views seem to be well placed to make this objection because these views openly reject the idea that self-expression requires motivational unification (Doris 2015: 162; Shoemaker 2015: 135-6; Sripada 2016: 1230). Sripada makes this explicit by introducing a “mosaic conception” of the real self, according to which “the contents of one’s self can be inconsistent and even in open conflict” (Ibid.), and a “narrow approach” to expression, according to which “a person’s action doesn’t need to express all of a person’s self; it can be anchored in small part—perhaps only a sliver” (1228). In response, recall that a central tenet of real self views is that not every psychic element is part of the agent’s real self; for Sripada in particular, the agent’s cares are its only constituents. Thus, it follows that, on his view, conflicts within the real self must necessarily involve opposing cares; similarly, even the sliver on which expression hinges must be one of the agent’s cares. Regarding inverse enkratics, however, the conflict occurs between something that is a care and something that, by hypothesis, isn’t, and it’s the latter—the agent’s inauthentic all-things-considered judgment—that gets expressed. This judgment actually is “external” to the agent’s real self[[7]](#footnote-7) and thus her action isn’t self-expressive at all.

 It might be thought that Shoemaker’s (2015) “ecumenical real self view” has the resources for countering this last claim, in that according to it the real self encompasses *both* the agent’s “authenticating” cares and her “authorizing” evaluative judgments. Actually, however, it’s hard to see how this proposal could successfully address inverse enkrasia cases. As Shoemaker concedes (120), evaluative judgments can be the product of false consciousness and, when this occurs, they, and the actions prompted by them, aren’t self-expressive. Presumably then, he must say that for an evaluative judgment and its resulting action to express the agent’s real self they must be sufficiently “internal”. But if what this means is that the judgment and the action must be connected to the agent’s authentic values and cares after all, then the latter take again center stage and thus my point that the judgments and actions of inverse enkratics (because detached from such values and cares) aren’t self-expressive would remain untouched. If, on the contrary, Shoemaker were to insist that actions prompted by objectionable evaluative judgments are sufficient grounds for blameworthiness regardless of their connection to the agent’s authentic values and cares, then he would have switched paradigms, adopting a sort of Scanlonian account[[8]](#footnote-8)—which assesses the moral implications of evaluative judgments without worrying about their psychic provenance—instead of vindicating the real self view.

 A third and crucial objection questions precisely the stipulation that the inverse enkratic’s all-things-considered judgment and subsequent action fail to reflect her values and cares. According to the objection, a more realistic description of my examples would depict these agents as caring *somewhat* about slavery, chastity, and selfishness.[[9]](#footnote-9) My response here is twofold. First, while it’s certainly possible that inverse enkratics hold conflicting values and cares, this isn’t necessarily the case. Another possibility is that their judgment is exclusively the product of exogenous factors rather than of something they truly—albeit partially—care about. Recall in this regard Sripada’s (2016: 1207) own stipulation that the puritanical man’s judgment “reflects the man’s acculturation” rather than one of his cares. It could be countered that one of the main things acculturation achieves is precisely to make people actually care for certain things through a process of norm internalization. This is true, but it’s also true that internalized norms needn’t have to be anchored in one of the agent’s cares for them to be motivationally efficacious. Repurposing a distinction introduced by Gibbard (1990), we can differentiate two ways in which a norm can be internalized: one is by *accepting* a norm, another is by *being in the grip of* a norm. In this context, we can say that an agent accepts a norm when its motivational efficacy comes from its fostering or aligning with one of her cares (which might themselves be the product of acculturation); by contrast, an agent is in the grip of a norm when the norm’s motivational efficacy bypasses her cares altogether and is purely the product of exogenous factors such as peer pressure, socialized habits, bizarre situational factors, etc. The puritanical man is in the grip of a norm directing him to avoid carnal pleasures because its motivational efficacy its entirely due to the man’s socialization rather than one of his cares; similarly, the enkratic Huck is in the grip of a norm directing him to hunt down fugitive slaves because its motivational efficacy runs contrary to his true cares and can be fully explained by the conventional morality of his time. To deny this distinction would be to imply that norm-based motivation is *necessarily* connected to cares, which is false. A person can regularly follow a certain norm out of habit only to realize at some point that all she has been doing is going through the motions without really caring about the prescribed behavior at all, as when someone who has rigorously followed a Kosher diet since childhood suddenly realizes that Kosher strictures mean nothing to her.

 My second response to the objection under consideration simply reminds the objector that a central tenet of anti-reflectivist real self views is that there is room for mistaken self-attribution and self-ignorance concerning one’s values and cares (Doris 2015: 161; Sripada 2016: 1212). Therefore, inverse enkratics of the sort I depicted above—agents whose evaluative judgments and their attendant motivations are wholly disconnected from what truly matters to them—are not only a realistic possibility but also one that anti-reflectivist real self theorists are firmly committed to.

A fourth objection concedes that agents in my examples might not care about slavery, chastity, or selfishness but argues that their actions are self-expressive nonetheless because they do reveal *different* cares of theirs. It can be argued, for instance, that in the case of Huck his action reveals a concern with property rights and, in the case of the young Randian, a concern with avoiding conformism (Arpaly 2015: 148). Sripada makes a related suggestion about the puritanical man, writing that “it might well turn out that there *is* a care of the man’s directed at pleasing his family and church community, which explains his having puritanical reactions” (2016: 1212). My response is that this move is illegitimate in the present dialectical context, since we can (almost) always redescribe a particular action so as to make it seem the product of one of the agent’s cares, including cases in which the real self theorist is inclined to deny that self-expression occurs. For instance, we saw above that Sripada claims that actions prompted by irresistible desires don’t express the agent’s real self; however, in the case of Frankfurt’s unwilling addict—which is Sripada’s prime example of a failure of self-expression—we could say that, while her action doesn’t express a care of hers related to consuming drugs, it does express her caring about, say, pleasure. Similarly with Doris’ (2015: 161) claim that a person who yields to coercion doesn’t express her values and thus isn’t responsible for her conduct: while it’s true that the coerced person doesn’t value, say, getting easy money, she does value remaining alive and this value of hers certainly gets expressed in her choosing to yield to, rather than to ignore, the robber’s demand to drive the getaway car. The lesson is that for real self theorists to be able to meaningfully draw the distinction between actions that are self-expressive and actions that aren’t, they have to pick what can be called a “focal care” of the agent’s on the basis of which this distinction can be made.

In the inverse enkrasia cases presented above, the focal care is identified by looking at the content of the agent’s explicit evaluative judgment; we then ask whether the agent’s action can be explained as the product of a care of hers that corresponds to this content. Following the original description of the cases plus the argument offered in response to the previous objection, I have claimed that we can legitimately give a negative answer to this question in the inverse enkrasia cases as well. These focal cares—concerning slavery, chastity, and selfishness—are the ones at stake in determining whether the inverse enkratic’s action expresses her real self or not. Whether *other* values and cares are expressed in her conduct is immaterial to this question.

Sripada might insist that the actions of inverse enkratics are self-expressive because they express the agent’s *flaws*, where the latter involve problematic cares and priorities that represent “only a small part of our self” but that nonetheless “remain fully part of who we are” (2016: 1226-7). However, the inverse enkratics in my examples don’t manifest flaws in this sense, because the non-focal cares that can be attributed to them—concerning property rights, pleasing one’s family, and avoiding conformism—aren’t problematic considered in themselves and so they aren’t flaws as Sripada understands them. Could it be that what is morally objectionable aren’t these cares considered in isolation but rather the *prioritization* of cares manifested in the inverse enkratic’s conduct, for instance the enkratic Huck’s objectionable prioritization of property rights over Jim’s well-being? I concede that *if* this prioritization of cares were behind the enkratic Huck’s action, then the latter would be self-expressive after all. However, in order to explain his action we needn’t suppose that Huck cares more about property rights than he does about Jim; on the contrary, it might even be that, deep down, Huck cares more about the latter than about the former (suppose that Huck has committed petty thefts now and then and so his allegiance to property rights is rather weak). Rather, what explains his action is simply his belief on the alleged rightness of turning Jim in. There is nothing mysterious about this stipulation, given the (imperfect) capacity human agents have to adjust their motivations to their normative beliefs. Thus, the internalized norm on whose grip he is in can itself motivate Huck without revealing a morally objectionable priority of cares actually held by him.

A final objection is that the intuition that inverse enkratics are blameworthy is unreliable, particularly in the cases of Huck and the young Randian both of whom are presented as thoroughly good people in the original description of the examples. In response, I concede that it would be nice to have empirical data about ordinary intuitions in cases of inverse enkrasia.[[10]](#footnote-10) Absent that, there are still solid theoretical grounds for thinking that inverse enkratics (including Huck and the Randian) really are blameworthy for their actions. As I pointed out above, by putting their objectionable judgments into practice they did evince ill will in this particular occasion—although perhaps less than wholehearted racists, puritans, and egoists would. So while they may be less blameworthy than the latter, they are blameworthy nonetheless. Moreover, this conclusion can be embraced by theorists who conceive of blame in very different ways, for instance as the appropriateness of reactive attitudes following violations of moral obligations one accepts (Wallace 1994); as the modification of relationships in response to attitudes that impair them (Scanlon 2008: Ch. 4); as a demand to defend one’s judgments (A. Smith 2005); or as a negative mark in the agent’s moral ledger (Zimmerman 1988). My inverse enkratics seem blameworthy under all of these different construals of what blame is.

 I thus conclude that none of the objections discussed above is persuasive and so my challenge to real self theorists stands: inverse enkrasia shows that self-expression isn’t a necessary condition on moral responsibility.

**5. Conclusion**

By appealing to the neglected phenomenon of inverse enkrasia I have shown that anti-reflectivist real self views are problematic, since they—like their reflectivist forerunners—are premised on the false assumption that moral responsibility requires the expression of the agent’s real self. At the same time, I have claimed that my argument leaves the quality of will paradigm unscathed, given that the inverse enkratics’ blameworthiness does depend on their evincing ill will. Although (as I pointed out in section 2) there is an understandable temptation to equate the agent’s evinced quality of will with the expression of her real self, the temptation should be resisted. Blame attributions are governed by a requirement of psychological accuracy: blame is warranted only if the agent did act for reasons showing lack of proper regard for others (Scanlon 2008: 180). Real self views go wrong in supposing that psychological accuracy also requires authenticity; that is, they mistakenly assume that a blameworthy quality of will must be indicative of who the agent really is—of what she truly values and cares about. I have shown that this isn’t the case. Therefore, those who want to defend the quality of will paradigm would do better to stop talking about real selves altogether.

**References**

Arpaly, N. and T. Schroeder (1999) “Praise, Blame, and the Whole Self.” *Philosophical*

 *Studies* 93 (2): 161-88.

Arpaly, N. (2015) “Huckleberry Finn Revisited: Inverse Akrasia and Moral Ignorance.” In

 R. Clarke, M. McKenna, and A. Smith (eds.) *The Nature of Moral Responsibility*, pp. 141-156. Oxford: Oxford University Press.

Doris, J. (2015) *Talking to Our Selves*. Oxford: Oxford University Press.

Faraci, D. and D. Shoemaker (2014) “Huck vs. JoJo: Moral Ignorance and the

 (A)symmetry of Praise and Blame.” *Oxford Studies in Experimental Philosophy,* Vol. 1, pp. 7–27. Oxford: Oxford University Press.

Frankfurt, H. (1971/1988) “Freedom of the Will and the Concept of a Person.” In *The*

 *Importance of What We Care About*, pp. 11-25. Cambridge: Cambridge University Press.

Gibbard, A. (1990) *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University

Press.

Hume, D. (1772/1978) *A Treatise of Human Nature*. Oxford: Clarendon.

Reis-Dennis, S. (2018) “Responsibility and the Shallow Self.” *Philosophical Studies* 175

 (2): 483–501.

Scanlon, T.M. (1998) *What We Owe to Each Other*. Cambridge, MA: Harvard University

Press.

Scanlon, T.M. (2008) *Moral Dimensions*. Cambridge, MA: Harvard University Press.

Shoemaker, D. (2015) “Ecumenical Attributability.” In R. Clarke, M. McKenna, and A.

Smith (eds.) *The Nature of Moral Responsibility*, pp. 115-40. Oxford: Oxford University Press.

Smith, A. (2005) “Responsibility for Attitudes: Activity and Passivity in Mental Life.” *Ethics* 115 (2): 236-71.

Smith, H. (2015) “Dual-Process Theory and Moral Responsibility.” In R. Clarke, M.

McKenna, and A. Smith (eds.) *The Nature of Moral Responsibility*, pp. 175-209.

Oxford: Oxford University Press.

Sripada, C. (2015). “Acting from the Gut: Responsibility without Awareness.” *Journal of*

 *Consciousness Studies* 22 (7-8): 37-48.

Sripada, C. (2016) “Self-Expression: A Deep Self Theory of Moral Responsibility.”

*Philosophical Studies* 173 (5): 1203-32.

Talbert M. (2012) “Moral Competence, Moral Blame, and Protest.” *Journal of Ethics* 16

(1): 89-109.

Taylor, C. (1991) *The Ethics of Authenticity*. Cambridge, MA: Harvard University Press.

Wallace, J. (1994) *Responsibility and the Moral Sentiments*. Cambridge, MA: Harvard

 University Press.

Watson, G. (1975/2004) “Free Agency.” In *Agency and Answerability*, pp. 13-32. Oxford:

 Oxford University Press.

Zimmerman, M. (1988) *An Essay on Moral Responsibility*. Totowa, NJ: Rowman &

 Littlefield.

1. Prominent defenders include Arpaly and Schroeder (1999), Shoemaker (2015), H. Smith (2015), Doris (2015), and Sripada (2016). These views stand in contrast to older *reflectivist real self* *views*, which were premised on the assumption that the agent’s real self coincides with the motives the person consciously endorses in deliberation (see Frankfurt 1971/1988; Watson 1975/2004). [↑](#footnote-ref-1)
2. Arpaly and Schroeder (1999) reject real self views and sketch instead a “whole self view”, but it’s clear that their target is *reflectivist* real self views. [↑](#footnote-ref-2)
3. See footnote 1 above. [↑](#footnote-ref-3)
4. Cf. Reis-Dennis (2018), who mistakenly equates the rejection of the real self view with the rejection of the quality of will paradigm. [↑](#footnote-ref-4)
5. See Taylor (1991) and Reis-Dennis (2018: sect. 2). [↑](#footnote-ref-5)
6. Conduct should be understood broadly as encompassing actions, omissions, emotional responses, patterns of (un)awareness, etc. [↑](#footnote-ref-6)
7. Sripada writes that “The care-based view allows that many evaluative judgments don’t bear any connection to the deep self, namely those that don’t bear the right dispositional tie to one’s cares” (2016: 1210, n. 12). [↑](#footnote-ref-7)
8. See in particular Scanlon (1998, 2008) and Talbert (2012). [↑](#footnote-ref-8)
9. See Doris (2015: 161) for this suggestion regarding Huck. [↑](#footnote-ref-9)
10. The closest we have is Faraci and Shoemaker’s (2014) studies about ordinary intuitions concerning the akratic version of Huck’s case. [↑](#footnote-ref-10)