

# Moral Worth and Consciousness: In Defense of a Value-Secured Reliability Theory

(penultimate draft; final version forthcoming in *Ergo*)

John W. Robison

*Indiana University, Bloomington*

**ABSTRACT:** What minimal role—if any—must consciousness of morally significant information play in an account of moral worth? According to one popular view, a right action is morally worthy only if the agent is conscious (in some sense) of the facts that make it right. I argue against this consciousness condition and close cousins of it. As I show, consciousness of such facts requires much more sophistication than writers typically suggest—this condition would bar from moral worth most ordinary, intuitively morally worthy agents. Moreover, I show that the attraction to this flavor of consciousness condition rests on mistaken assumptions about what is required for a right act to be non-accidentally right and attributable to the agent. Drawing some lessons from the discussion, I defend a Value-Secured Reliability Theory of Moral Worth and show how a minimal yet indispensable role for consciousness falls out from it. On this independently plausible theory, an action can be morally worthy even when the agent is unaware of the right-making features of her action.

## 1. Introduction

Huckleberry Finn is considering whether to continue helping Jim escape from slavery.<sup>1</sup> Consciously, Huck believes that it is morally wrong to help a slave escape—his conscious belief reflects the popular and mistaken norms of his society. A ripe opportunity to turn in Jim presents itself. All the while believing that he may go to hell for doing what is wrong, Huck decides to continue helping Jim escape, even though he has no clear story he can tell himself or others about *why* he helps Jim escape.

Though Huck is unaware of the moral significance of his action, many find it intuitive that his action is not only morally right but morally worthy: non-accidentally right and attributable to *him*.<sup>2</sup> If this is correct (and I believe that it is), then Huck-like cases raise a puzzle about the relationship

---

<sup>1</sup> Bennett (1974) is widely taken to be the first to bring Huck's case to the attention of moral psychologists. Arpaly and Schroeder (1999) reignited interest in the case.

<sup>2</sup> A quick note on terminology. Throughout this paper (and following the literature), I will use "S's action has moral worth" interchangeably with "S's action is praiseworthy" and "S is morally responsible for her right/good action." I further clarify the relevant terminology and scope of this project at the beginning of Section 2.

between moral worth and consciousness.<sup>3</sup> On one hand, such cases suggest that there is some sense in which moral worth does not require awareness of the moral significance of one's action. On the other hand, consciousness must play *some* role in securing moral worth: Huck's behavior was not morally worthy if he was in a trance and altogether cognitively disconnected from the morally significant features of his situation. So, what minimal role *does* consciousness play in securing moral worth?

According to one popular view, a right action is attributable to the agent and non-accidentally right as moral worth requires only when the agent is conscious of the facts that make it right. Intuitive as this may be, I argue that this cannot be the minimal consciousness condition on moral worth. As I show, consciousness of such facts requires *much* more sophistication than writers typically suggest—this condition would bar from moral worth most ordinary, intuitively morally worthy agents. Moreover, I show that the attraction to this general variety of consciousness condition rests on mistaken assumptions about what is required for a right act to be non-accidentally right and attributable to the agent. We want to know: what is the most minimal sense in which an agent must be conscious of the moral significance of her action if the action is to be non-accidentally tied to the right and attributable to the agent as moral worth requires? This paper shows what is mistaken about the popular response, and it offers an account of moral worth from which a minimal yet indispensable role for consciousness falls out.

Here is how things proceed. Section 2 clarifies our question and specifies some desiderata for a minimal consciousness condition. In Section 3, I show, first, that the depth of disagreement between the Anti-Consciousness Camp (those theorists who actively seek to downplay or eliminate any role

---

<sup>3</sup> This paper largely proceeds on the assumption that it *is* a datum that there is some version of Huck's story on which his action is morally worthy. However, for those who are skeptical that Huck's action is morally worthy, the account of moral worth I offer in Section 4 can supply an argument that, on a plausible account that secures everything we might want from an account of moral worth, Huck-like actions can be morally worthy.

for consciousness) and the Pro-Consciousness Camp (those theorists who seek to emphasize its role) has been exaggerated.<sup>4</sup> Across *both* camps, there is widespread commitment to the aforementioned view that an action is morally worthy only when the agent is conscious of the facts that make it right. The section continues by suggesting that (at least on some natural interpretations) this is not a necessary condition on moral worth: the condition sets unreasonably demanding and excessively restrictive standards for moral worth. It then argues that the attraction to this general variety of consciousness condition rests on false assumptions about the requirements for non-accidentality and attributability. Drawing some lessons from Section 3, Section 4 develops and defends what I call the “Value-Secured Reliability Theory of Moral Worth.” On this view, an agent’s action has moral worth just to the extent that its production is explained by her *value-secured reliable tie to the right*, a reliable tie to the right that is secured through the influence her person-level values have (perhaps unreflectively) on her patterns of informational access and processing.<sup>5</sup> As I show, a minimal yet indispensable role for consciousness falls out from this account, for consciousness is the integral vehicle through which an agent’s action can be explained by a value-secured reliable tie to the right. Moral worth does not require that an agent be conscious of *the fact(s) in virtue of which her action is right*—it requires just that

---

<sup>4</sup> See Arpaly (2003, 2015a, 2015b), Arpaly and Schroeder (1999, 2013), and Sher (2009) for discussions from the *Anti-Consciousness Camp*. Levy (2011a, 2011b, 2014) is among the key representatives of the *Pro-Consciousness Camp*.

<sup>5</sup> My notion of *value-secured reliability* is in important respects indebted to Ernest Sosa’s work in virtue epistemology (2007, 2015). Sosa has long argued in some form or other that <reliability secured by something attributable to the agent> marks an important category in epistemology. Irrespective of whether Sosa is correct about the epistemology, I believe that this broad category is central to moral worth. This being said, there are important differences between me and Sosa on what makes reliability attributable to *the agent* (or something for which *the agent* is responsible), and these differences matter a great deal for the debate about moral worth and consciousness. Roughly, Sosa locates attributability to the agent in an agent’s “second-order awareness” of his own reliability (2015: 79). Such “reflective competence” is the key ingredient underwriting Sosa-ian concepts such as “reflective aptness” and “aptness full well” (76): when an agent achieves the status of “reflective aptness,” her belief is reliably produced *and* sustained in virtue of her appreciation of the fact that it is reliably produced. It is *the agent’s* appreciation and active endorsement of her own reliability (or aptness) that makes it genuinely attributable to *her as an agent* (it is only in such cases that one’s “rational nature is most fully manifest” (2015: 51)). I reject any such meta-awareness or meta-competence condition for attributability. On my Value-Secured Reliability Theory, reliability is attributable to the agent simply to the extent that it has been secured by the agent’s values having shaped the inputs and outputs of the relevant cognitive and decision-making mechanisms. One’s values can have this shaping effect completely unreflectively and without the agent’s having any metarepresentational grasp of her own reliability. To the extent that the reliability is secured by influence from *person-level values* (as opposed to some God-hand), the reliability is properly attributable to *the agent*.

certain information was accessed and processed to produce the right action *because* of the agent's value-secured reliable tie to the right. Depending upon the agent's background values, the strength of those values, and the influence her values have had on shaping her information accessing and processing mechanisms, she can perform morally worthy actions even when conscious of nothing more robust than quite minimal sensory cues. This independently plausible account resolves our puzzle about the relationship between moral worth and consciousness: it secures both non-accidentality (*reliability*) and attributability (*value-secured*), and it gives consciousness a clear and well-motivated role.

Let us clarify the central question of this paper.

## **2. Clarification, Disambiguation, and Desiderata on a Minimal Consciousness Condition**

Before turning to what others have said about our question, let me make two clarificatory remarks. First, the focus of this project is on just the *positive* side of moral responsibility: it asks what cognitive or epistemic conditions are required for a *right* action to be morally worthy (or praiseworthy). Given that there are some compelling reasons to believe that the conditions for praiseworthiness and blameworthiness may be asymmetrical (see Wolf 1980, 1987), it may well be that blameworthy action and praiseworthy action have different cognitive or epistemic requirements. The focus here is on just the success case: what must an agent be aware of if his right action is to be non-accidentally right and attributable to him in the way that would make him praiseworthy for it?

A second, terminological point: as already illustrated above, I shall be using “morally worthy,” “praiseworthy,” and “responsible for one's right action” interchangeably. This follows Nomy Arpaly, who, on writing about Huck Finn, indicates that she will “speak interchangeably of a ‘morally praiseworthy action’ and ‘an action with positive moral worth’” (2003: 69). Those responding to Arpaly, like Neil Levy, have correctly taken her to be targeting *praiseworthiness* and *responsibility for right action*—referencing the very chapter where Arpaly indicates that she will use “morally worthy” and

“praiseworthy” interchangeably, Levy writes that “when Arpaly argues that Huck can be responsible for his actions despite being incapable of bringing to consciousness the reasons to which he, nevertheless responds, it is clearly awareness that she denies is needed to ground moral responsibility” (2014: 35). While I will tend to use the term ‘morally worthy,’ the term could be replaced with ‘praiseworthy’ in every instance.

Our central question is: what is the most minimal sense in which an agent must be conscious of the moral significance of her action if the action is to be morally worthy? The literature tends to approach this question by asking *whether* moral worth requires that one be conscious of the moral significance of her action. In this section, I suggest that this approach has been unhelpful, and I clarify what it is that we are looking for when we are looking for a minimal consciousness condition on moral worth.

In *Consciousness and Moral Responsibility*, Neil Levy explicitly labels Nomy Arpaly and George Sher opponents of his view that consciousness of the moral significance of one’s action is necessary for moral worth (2014: 77). Moreover, Arpaly explicitly identifies herself as an opponent of Levy’s view in her review of his book (2015b: 829). This might suggest that there is a well-defined dispute about whether moral worth requires consciousness of the moral significance of one’s action, with the Anti-Consciousness Camp on one side and the Pro-Consciousness Camp on the other. However, the turn of phrase “S is conscious of his action’s moral significance” can pick out a variety of substantively different cognitive relations between an agent and some bit(s) of information. Once we disambiguate this turn of phrase, we will see that it is unclear that the debate between the Pro-Consciousness Camp and the Anti-Consciousness Camp is as well-defined as it might appear.

What might be meant by the claim or denial that moral worth requires “consciousness of the moral significance” of one’s action? Some passages from the Anti-Consciousness Camp leave things mysterious. Arpaly writes that Huck is morally worthy but “not capable of bringing to consciousness

his *nonconscious awareness*” of Jim’s humanity (2003: 77). But what precisely is this relation of *nonconscious awareness*? “Conscious” and “aware” are often used synonymously in this context, so it is unclear what specific cognitive relation Arpaly has in mind. George Sher, another card-carrying member from the Anti-Consciousness Camp, similarly leaves things mysterious when he writes that “agents can satisfy responsibility’s epistemic condition by accurately but unconsciously processing the *information to which they have access*” (2009: 143) and that satisfying this condition requires just that an agent have “made enough *cognitive contact*” with the evidence for an action’s moral rightness (143). Given that a common referent for “conscious” is Ned Block’s (1995) *access conscious* (according to which information is conscious just when an agent has the right kind of *access* to it), Sher invites confusion through his liberal use of “information to which an agent has access” while defending an Anti-Consciousness position. When situated in a debate about what bearing—if any—consciousness has upon moral worth, passages like these obscure matters.

To understand some claim or denial that moral worth requires that one be “conscious of the moral significance” of one’s action, there are two things we need to know. First, we need to know what *cognitive relation* between an agent and some target information is under discussion. Is the claim about whether an agent must be *access conscious* of certain information, or about whether an agent must have *consciously and effortfully deliberated upon* certain information, or about whether certain information must be *readily available for report*, or...? Second, we need to know what *kind(s) of information* of which an agent can be conscious is under discussion. Whatever the relevant cognitive relation should be, is the claim about whether an agent must stand in that relation to facts about the *deontic status of his action*, or to some *non-deontic moral facts about his situation under explicitly moral concepts* (e.g., “Jim deserves respect” or “Jim is being treated unfairly”), or to the *non-moral facts upon which some moral reasons supervene* (e.g., “Jim is in pain” or “Jim is not living the life he wishes to live”), or...?

Given the numerous candidates for both the cognitive relation and the kind(s) of information identified above, there are several things a writer may have in mind in claiming or denying that moral worth requires consciousness of the moral significance of one's action. Moreover, some versions of this claim are much more demanding than others. Information can pop into mind and be access conscious without the agent necessarily *consciously deliberating* upon it. And, an agent can be aware of some *non-deontic moral facts about his situation under explicitly moral concepts* without being aware of the *deontic status of his action* (Huck might be aware that Jim is deprived of respect while being unaware that helping him is morally required). Any minimal consciousness condition should identify the most minimal combination of cognitive relation and kind(s) of information that is necessary for moral worth (and sufficient as far as consciousness is concerned, bracketing any other potential conditions on moral worth).

To be motivated, a minimal consciousness condition should play some nontrivial explanatory role in securing two desiderata on moral worth: namely, a non-accidental tie to the right and attributability to the person. After all, concern for these features is, I take it, what makes a consciousness condition of any sort attractive in the first place. When I inadvertently donate to some charity while unaware of what button I am pressing, my action is not morally worthy because it lacks the non-accidental tie to the right. Moreover, when I am not appropriately aware of my circumstances, the morally significant features of my action are not expressive of my person-level beliefs and desires—my unwitting donation to the charity does not involve the right connection to my person-level attitudes that moral worth requires. So, satisfying the minimal consciousness condition ought to play some significant explanatory role in securing non-accidentality and attributability.

In evaluating some purported minimal consciousness condition, we must ask two questions. First, we should ask: does satisfying the condition help secure non-accidentality and attributability? If the answer is “no,” then we have some reason to reject it. Second, we should ask: can a consciousness

condition *more* minimal than the proposed one secure non-accidentality and attributability at least as well as the purported minimal consciousness condition (is the purported condition really a *necessary* condition)? If there *is* some more minimal condition, then the proposed one cannot be a necessary condition on moral worth, and we should reject it.

Having clarified the shape and desiderata of an adequate minimal consciousness condition, let us locate and evaluate a popular answer that shows up across *both* the Anti-Consciousness Camp and Pro-Consciousness Camp.

### **3. Consciousness of the Right-Making Facts**

According to this popular answer, a right action is morally worthy only when the agent is in some sense conscious of the facts that make it right. In this section, I first locate this position across both the Pro-Consciousness Camp and (more surprisingly) the Anti-Consciousness Camp. The aim of the section is, first, to cast doubt on this popular purported necessary condition on moral worth. I argue that, on various interpretations of “conscious of the facts that make one’s action right,” the conditions for moral worth would be excessively restrictive if they required such consciousness. This gives us good reason to doubt that the condition is necessary for moral worth: the condition would rule out lots of intuitively morally worthy cases. Furthermore, after considering variations of this condition that might be less restrictive, the section continues by undercutting the general motivations for any version of a minimal consciousness condition that involves consciousness (in some sense) of the facts in virtue of which one’s act is right (the case begins in this section, but the full case is made in conjunction with Section 4). Consciousness conditions of this general sort are motivated by the thought that such consciousness is somehow intimately related to the agent’s right action being *non-accidentally* right in a way that is *attributable* to the agent. But proponents of such conditions operate with mistaken assumptions about what is required for non-accidentality and attributability. I point to



examples that illustrate that satisfying such consciousness conditions is certainly not *sufficient* for making a right action non-accidentally right in an agent-attributable way, and I suggest that, when we see what is missing in these examples, we will see that the further resources we need to call upon are—at least in principle—capable of explaining moral worth without involving the kind of consciousness conditions considered. It is then in Section 4 that I offer a theory of moral worth that shows—in light of how our informational access and processing mechanisms actually work—that morally worthy actions do not require satisfying any of these consciousness conditions. Consciousness, I suggest, plays an integral but different and much more limited role in a theory of moral worth.

Let us begin by focusing on the role given to consciousness in the work of Sher, Arpaly and Schroeder from the Anti-Consciousness Camp. It is clear that these writers reject any minimal consciousness condition with *deontic status* as the information of which one must be conscious. Arpaly writes that “for an agent to be praiseworthy for an action, it is *not required* that she believe that what she does is right” (2015a: 145), and Sher makes similar suggestions (2009: 143). Moreover, if we take as a datum that Huck-like actions can be morally worthy, then the minimal consciousness condition cannot have *deontic status* as the target information, as Huck is, by hypothesis, not conscious (in any sense) of this information.

It is also clear that these writers reject any minimal consciousness condition with *deliberation* as the relevant cognitive relation. In describing Huck’s morally worthy action, Arpaly writes that Huck “constantly perceives data (*never deliberated upon*) that amount to the message that Jim is a person, just like him” (2003: 76). None of these writers makes moral worth contingent upon the agent’s having consciously or effortfully weighed moral reasons.

While these writers clearly deny that *deliberation* is required for moral worth, things are less clear concerning *awareness* (understood as *wide availability of the relevant information to mechanisms in the agent*).<sup>6</sup> In reviewing Levy's (2014) *Moral Responsibility and Consciousness*, Arpaly writes that "some philosophers, including yours truly, have argued against the Consciousness Thesis," the thesis that "to be morally responsible for an action, one needs to be aware of those features of the action that make it good or bad" (2015b: 829). This sounds as though Arpaly is denying that moral worth requires that an agent be aware of those features of his action that make it good. However, in this review, the cases Arpaly eventually cites as counterexamples are just about *deliberation*: they involve a jazz musician who is praiseworthy for his improvisation even though "he has no time for conscious deliberation" (2015b: 830) and a praiseworthy witty conversationalist who "does not deliberate before every funny comment" (831). Whether an agent has *deliberated* upon some information is different from whether he was *aware* of or had *access* to that information without having effortfully deliberated upon it.

What is more, Arpaly's descriptions of Huck suggest that awareness (albeit, not of *deontic status*) *does* play a crucial role in securing moral worth on her account. Arpaly writes (all italicizing is my own) that "while Huckleberry does not conceptualize his realization, it is [an] *awareness* of Jim's *humanity* that causes him to become emotionally incapable of turning Jim in" (2003: 10). Huck, she writes, "constantly *perceives data* (never deliberated upon) that amount to the message that Jim is a *person*, just like him" (76). And, with Schroeder, Arpaly offers the following extended interpretation of Huck:

Different interpretations of the novel are possible, but one possible interpretation (not unrealistic, and one we favor) is that Huckleberry is motivated to not turn in Jim because Huckleberry intrinsically desires what is right or good via the relevant concepts, the ones that would be identified by a correct normative moral theory, and *sees that this end will be promoted by Jim's escape ...* . Huckleberry *sees* that Jim's life as a slave, separated from family against his will, always forced to do what another says, and never compensated for his efforts, is lived in the absence of the *respect* he intrinsically desires everyone to enjoy. On one interpretation of the novel, *these things might all have come clearly to Huckleberry's mind*, and have weighed heavily with him

---

<sup>6</sup> Awareness in this sense roughly picks out what Ned Block (1995) refers to as "access consciousness" and what others refer to as information being "globally broadcast" in the Global Workspace model of consciousness (see Baars, 1988).

emotionally because of his strong intrinsic desire that everyone be treated with *respect*. And this might well have happened without Huckleberry ever concluding that Jim's escape from slavery would be right or good—might have happened while Huckleberry self-consciously concluded that what is right or good is to return Jim to slavery (2014: 178).

A few things are worth noting. First, *awareness* is clearly part of the story underwriting Huck's moral worth (the passages reference Huck's "awareness," the "data" he "perceives" or "sees," and the things that have "come clearly to Huckleberry's mind"). More than this, Huck is described as being aware of moral facts about his situation *under explicitly moral concepts*. He sees that Jim's life is lived in the absence of the *respect* he deserves, and he is aware of Jim's *humanity* or that he is a *person* (it is clear that "humanity" and "person" pick out the forensic categories and not the biological categories). On this reading, Huck *is* aware of moral features of his situation under explicitly moral concepts, he just does not *deliberate* upon these facts, and he never arrives at a belief that, all things considered, it is morally right to help Jim. So, the language used to describe Huck suggests that Arpaly and Schroeder are committed not only to a role for awareness in moral worth but to a role for awareness of (non-deontic) moral facts *under explicitly moral concepts*.

Now, I suspect that Arpaly and Schroeder would not endorse such an explicit commitment to the significance of an agent's awareness of their situation under *explicitly moral concepts*. Huck, Arpaly writes, acts "for the reasons that make [his action] right," but he "does not *know* that they are moral reasons" (2015a: 143). So, perhaps when they write that the facts about Jim's "humanity/personhood" or the "respect he deserves" come clearly to Huck's mind, this is just shorthand for "the nonmoral facts upon which respect/personhood supervenes." In that case, the view on offer is that moral worth requires that one be aware of the morally relevant nonmoral facts upon which the moral rightness of one's action supervenes. Even if *this* is the view (rather than the view that requires explicitly moral conceptualizations), it turns out—somewhat surprisingly—that these representative writers from the Anti-Consciousness Camp give an important role to awareness after all.

George Sher, another member of the Anti-Consciousness Camp, appears similarly implicitly committed to a role for awareness. Sher writes that we can correctly capture the “crucial epistemic linkage” between an agent and moral reasons that moral worth requires only by “removing that linkage from the conscious realm” (2009: 143). This may sound like a rejection of *any* consciousness condition. However, the ways that Sher gestures toward that “crucial epistemic linkage” suggest that—like Arpaly—he intends to reject only consciousness conditions that involve deliberation or information about the deontic status of an action. Consider the following passages (italicizing is my own):

...[A]gents can satisfy responsibility’s epistemic condition by accurately but unconsciously processing the information *to which they have access* (2009: 143).

When someone performs an act in a way that satisfies ... any other conditions for responsibility that are independent of the epistemic condition, he is responsible for his act’s morally or prudentially relevant feature if ... he is unaware that the act is right or prudent despite having made enough *cognitive contact* with the evidence for its rightness or prudence to enable him to perform the act on that basis (143).

What precise combination of cognitive relation and type of information underlies the “crucial epistemic linkage” required for moral worth, according to these passages? Here is a natural reading. Certain facts count as evidence for the rightness of an act. Presumably, these are the nonmoral facts upon which the moral rightness of the action supervenes. When an agent has “access” to these facts (or has made “enough cognitive contact” with them), the information is access conscious and, thus, made widely available for nonconscious processing by various mechanisms in the agent. Of course, the information may be access conscious without the agent’s having any awareness of *how* her awareness of these facts will influence her subsequent behavior.<sup>7</sup> But, this nonconscious processing of the nonmoral facts upon which the moral facts supervene allows the agent to perform the morally right action *on the basis* of those reasons that make it right. If something like this is Sher’s view (he never makes precise what cognitive relation he has in mind), then he, too, is ultimately committed to

---

<sup>7</sup> Sher’s way of putting this—that *the agent* has access to the information—is misleading. Better to say that sub-personal mechanisms in the agent have access to the information.

the view that moral worth requires that one be aware of the nonmoral facts upon which the moral rightness of one's action supervenes.

So, key representatives from the Anti-Consciousness Camp are committed to a necessary, nontrivial role for consciousness in a theory of moral worth. Is the view on offer—the view that moral worth requires that the agent be aware of the nonmoral facts upon which the moral rightness of his action supervenes (or that he be aware of “the reasons that make it right” [Arpaly 2015a: 143])—a plausible minimal consciousness condition on moral worth? I suggest that it is not. The problem is that this condition is excessively restrictive—it sets cognitive requirements that Huck and other ordinary intuitively morally worthy agents cannot reasonably be expected to meet.

Note that *what makes* Huck's action right is some complex conjunction of many facts: “Helping Jim escape 1) helps a friend 2) who is not about to cause a bunch of harm 3) and who will otherwise have his freedom impinged upon by another agent 4) to whom he never non-coercedly gave his consent to be treated that way, where 5) ...”. The concern is that we cannot expect Huck to be aware of all of this—he cannot be aware of *the nonmoral facts in virtue of which his action is right*. That set of facts is too complex for Huck (as it is for the rest of us).

One way to reinforce this point is to note that, if Huck is really to be aware of *the fact in virtue of which his action is right*, it will not do for him to simply be simultaneously aware that (1) obtains, that (2) obtains, that (3) obtains, and so on. Awareness of each of these facts (individually) does not yet make Huck aware of *the fact in virtue of which his action is right*. To be aware of *that* fact, Huck needs to be aware of the *union* or *conjunction* of the facts above: he needs to be aware that “[1) & (2) & (3) & (4) & ...] obtains.” That long conjunction is the fact in virtue of which his action is right. Were Huck aware of and responsive to *that* fact, this may seem relevant to establishing that his right act is morally worthy (though, the claim that it “seems relevant” is quite distinct from the claim that it is *necessary*). But awareness of this long conjunctive fact is surely out of reach for someone with Huck's cognitive

abilities (as it is out of reach for any of us). This gives us some reason to doubt that moral worth requires *awareness of the nonmoral facts upon which the rightness of one's act supervenes* (at least, if we are not to be skeptics about moral worth).<sup>8</sup>

One might argue that my criticism of this potential necessary condition on moral worth has been a bit quick. Is it really true, in general, that *awareness of the nonmoral facts upon which the rightness of one's act supervenes* involves awareness of some complex, unwieldy conjunctive fact? Consider the rightness of taking care of a sick child. Couldn't one be aware of the nonmoral facts in virtue of which this is right by being aware of a small handful of facts (e.g., “this child is vulnerable and I can help her”)?<sup>9</sup>

If this objection to my criticism seems plausible, it may be because we are conflating *the salient reason(s) we would most naturally cite in defense of an action* with *the nonmoral facts upon which the rightness of some act supervenes*. It would not be right to keep sitting there caring for the sick child if you see two children drowning in the pond right outside. When we say that “taking care of the sick child is right because she is vulnerable and I can help her,” this is just a loose way of speaking—we are implicitly assuming that a whole mess of conditions is in place (and that a mess of other conditions is *not*). If moral worth required awareness of the nonmoral facts in virtue of which the rightness of one's act supervenes, then it would require awareness of that whole conjunctive mess about which conditions are or are not in place, and this asks too much of ordinary agents.

Thus far, I have argued that *awareness of the facts in virtue of which one's action is right* is not necessary for moral worth—that condition would make moral worth far too difficult to come by. We might tinker with the condition to make it more easily satisfiable. Suppose we say that moral worth requires,

---

<sup>8</sup> Note that it will be of no help to suggest that Huck can simply be aware of his situation under familiar Utilitarian or Kantian concepts. It is implausible to suggest that Huck and other non-ethicists are morally worthy only when aware that “my act is maximizing utility” or that “I can rationally will that the maxim of my action become universal law.”

<sup>9</sup> I thank an anonymous referee at *Ergo* for pressing me on this line of argument.

instead, awareness of the nonmoral facts upon which the relevant *pro tanto* moral reasons supervene (Arpaly and Schroeder seem torn between the former and the latter view; for an articulation of the latter, see 2014: 166).

Now, this condition is a bit vague (which are the “relevant” *pro tanto* moral reasons?). But, at least the condition seems more readily satisfiable than the previous one. Still, there is a problem for this view: the connection between *the nonmoral facts upon which some relevant pro tanto moral reasons supervene* and *the rightness of one’s act* is so weak that it is hard to see how awareness of the former could play any significant role in explaining why the performance of the right act was non-accidental and properly attributable to the agent. Since concern for non-accidentality and attributability is what motivates any version of a consciousness condition in the first place, this fact should give us pause. Certainly, such awareness is not *sufficient* for making a right act morally worthy. And once we see what other factors outside of such awareness need to be called upon to transform a right act into a morally worthy act, we will begin to see why awareness of the nonmoral facts upon which relevant *pro tanto* moral reasons supervene plays no necessary role in explaining what makes a right act morally worthy. Consider.

Stipulate that *helping a friend* is *pro tanto* morally important. Is Huck’s action morally worthy when, aware of the fact that he can help a friend, he helps Jim? Perhaps, but perhaps not. It had better not be the case that Huck, when aware that he can help a friend, will help the friend *no matter what the cause*. If Huck’s awareness of the fact that he can help a friend would, itself, cause him to help a friend execute a broad range of morally wrong actions (steal, physically harm people for fun, et cetera), then the fact that Huck is aware of and motivated by this fact certainly does not make his action morally worthy. In such a case, Huck’s action is not non-accidentally tied to the right (and, relatedly, the

successful performance of the *right* action is not properly attributable to *him*—his “success” is not well explained by his person-level values).<sup>10</sup>

When Huck satisfies this consciousness condition, his right action can still fail to be non-accidentally right in a way that is attributable to him—it is this failure that seems to explain why his action is not morally worthy in this case. What is missing? Suppose we revise the case a bit. Suppose Huck’s awareness that he can help a friend does not automatically make him help a friend, for he has a set of background values and desires that makes him reliably disposed to become aware of and appropriately responsive to nonmoral facts upon which the relevant *pro tanto* moral reasons supervene (at least in situations like the one he is in currently). When his action is explained by this reliable disposition, Huck’s action plausibly *is* morally worthy when he helps Jim while aware that he can help a friend: his act would be non-accidentally right in a way that is attributable to him (for, his values and desires explain his reliable disposition). But, now, it seems that this fact—the fact that his action is explained by this set of background values and desires that is reliably conducive to right action in the situation—*entirely* explains why his action is morally worthy.<sup>11</sup> *That* fact—*itself*—explains how his act is non-accidentally right in a way that is attributable to him. If it turned out that, even without awareness of the nonmoral facts upon which relevant *pro tanto* moral reasons supervene, an agent’s right action could still be explained by his having a set of background values and desires that is reliably conducive to the right (in situations relevantly like the actual one), I believe that we would still consider such a right action with such an explanation morally worthy. Awareness of *something* may (as an empirical fact) be necessary for one’s set of background values and desires to explain one’s response to one’s environment in a given situation. However, I will suggest in a moment that awareness of

---

<sup>10</sup> Moreover, it is worth noting that if we try to fix this problem by requiring that the agent be aware of *more* of the nonmoral facts upon which *more* of the relevant *pro tanto* moral reasons supervene, we risk running back into the cognitive overload worry from the previous version of a consciousness requirement.

<sup>11</sup> This point is more fully developed in Section 4.



information much less robust than *the nonmoral facts upon which the relevant pro tanto moral reasons supervene* will do. If all of this is correct, then awareness of the nonmoral facts upon which the relevant *pro tanto* moral reasons supervene is not necessary for moral worth.

Some might think that, since the connection between awareness of *pro tanto* moral reasons and the performance of a right action is so weak, what we ought to do is come up with a different potential minimal consciousness condition, the satisfaction of which would more plausibly play a role in explaining why some right action is non-accidentally right in a way that is attributable to the agent. We might think: to play a plausible, necessary role in making a right act morally worthy without setting unreachable cognitive requirements, the minimal consciousness condition must require that an agent be aware of at least *some* of the (most?) morally relevant features of his situation under *explicitly moral concepts*—it was a mistake to think that a consciousness condition without such moral conceptualization could accomplish what we would want from a minimal consciousness condition. This revision might seem to give us a condition that could plausibly play some necessary explanatory role in making a right act *non-accidentally* right: after all, were Huck to satisfy this condition, he would actually (and aptly) conceive of his right action under some of the relevant moral concepts. Moreover, the revision might solve the over-sophistication worry: rather than requiring that Huck be aware of that whole conjunctive mess of nonmoral facts upon which the relevant moral facts supervene, we might just require that Huck be aware of his moral situation under some more tractable moral conceptualization (e.g., aware that *Jim is not given the respect he deserves*). This would be disappointing for moral psychologists who aim to downplay the significance of an agent's moral conceptualizations (Arpaly and Sher included), but so be it.

Now, this modified consciousness condition is, again, a bit vague (*which* morally relevant features of one's situation must one be aware of under explicitly moral concepts?). But, set this worry aside. What I will show is, first, that there are cases where satisfying this consciousness condition is

still not—itself—sufficient for making one’s right act morally worthy. And I will then show, again, that once we see what is missing in these cases, we will see that the consciousness condition under consideration plays no necessary role in explaining what makes a right act morally worthy. Let us consider a case that illustrates how satisfying this condition does not suffice for making a right act non-accidentally right.

Suppose Soprano is aware of his circumstances under the relevant explicitly moral concepts. He is aware 1) that he made a promise to Baritone to kill Tenor, 2) that it is *pro tanto* good to keep one’s promises, 3) that killing is *pro tanto* wrong, 4) that loyalty (of some sort) is a virtue (of some sort), and 5) that Tenor has been loyal to him.<sup>12</sup> Suppose, in light of all this, that Soprano does the morally right thing: he does not kill Tenor. Soprano, thus, does the morally right thing *and* he is aware of the morally relevant features of his situation under explicitly moral concepts—he satisfies the minimal consciousness condition under consideration. Is his act morally worthy?

There is an obvious version of the story in which it is not. Suppose that, in the nearby possible world where Tenor had not been “loyal,” Soprano would not have hesitated to keep his promise to kill Tenor. The truth of this counterfactual makes it the case that Soprano’s right action in the actual world is not morally worthy, even if, in the actual world, he was aware of the right-making features of his action and acted partly on the basis of them. Indeed, this counterfactual would undercut Soprano’s moral worth even if he were aware, in the actual world, *that his act is morally right*. Soprano’s right action fails to be properly non-accidental. Awareness of the morally relevant features of one’s situation (whether conceptualized under explicitly moral concepts or not) is insufficient to transform a right act into a morally worthy act.<sup>13</sup>

---

<sup>12</sup> If the reader does not like “keeping promises” or “killing” as *pro tanto* rights and wrongs, feel free to substitute favorite *pro tanto* rights and wrongs.

<sup>13</sup> Following Paulina Sliwa (2016, see 401 in particular), one might argue that the problem with Soprano above is that, even if he *believes* or is *aware* that his action is morally right, he does not *know* that it is morally right (after all, he would change his mind in a very nearby possible world). Sliwa argues that moral worth requires that the agent *know* that his action is morally right. *Knowledge*—as opposed to mere awareness—has a counterfactual robustness. One knows that p only if one’s

What is missing in Soprano's case such that his act is not morally worthy? We can raise the same general point we raised earlier in response to the previous consciousness condition. Even when one is aware of morally relevant facts, one's right action may still fail to have been produced by a set of values and desires that is reliably conducive to right action in the situation. When a right action *is* explained by such a set of background values and desires, it is plausible to think that the action—produced in awareness of various relevant moral facts—*is* morally worthy. But, as we saw before, it now seems that the fact that the right action is explained by a set of background values and desires that is reliably conducive to the production of right action in such situations is—*itself*—capable of explaining *entirely* why this right action is morally worthy. That fact explains why the act is non-accidentally right in a way that is attributable to him. Perhaps if awareness of various relevant moral facts (morally conceptualized) from the agent's situation were *necessary* for the agent's right action to be explained by a set of background values and desires that is reliably conducive to the right in such situations, *then* we should think that such awareness is necessary for moral worth. But I argue in Section 4 that the facts about how our background values and desires can shape our informational access and processing suggest otherwise.

In the following section, I defend a new theory of moral worth and show how, on this independently plausible theory, a right act can be non-accidentally right in a way that is attributable to the agent—and, so, morally worthy—even when the agent is not aware of anything like the facts in virtue of which his action is right. But before I turn to this theory, one more point is worth exploring

---

belief in *p* is at least somewhat secure. We might think that Soprano's action is non-accidentally right only if he *knows* that his action is morally right.

Now, it *might* be sufficient for moral worth if I both know that my action is right *and* my action is caused in the right way by my moral knowledge. Specifying the “caused in the right way” clause would require some work, but I'll assume that this can be done in a way that would make for a plausible sufficient condition on moral worth. Still, even if this would make for a sufficient condition on moral worth, there is no reason to think that it is a *necessary* condition on moral worth. We should think this is a necessary condition on moral worth only if moral knowledge is the *only* way to secure non-accidentality and attributability, and I'll argue in Section 4 that we can secure these through the satisfaction of a more minimal consciousness condition (one that does not involve moral knowledge).

so that the general variety of minimal consciousness condition under consideration is given a fair hearing. One might think: even if an action could—in virtue of being explained by some reliable disposition to perform right action in the circumstances—be non-accidentally right without the agent’s meeting the aforementioned consciousness conditions, such an action could not be non-accidentally right *in a way that is attributable to the agent*. After all, goes the thought, how can we credit *the agent* with the performance of a non-accidentally right action if he is not even aware of the facts that make his action right? How, without such awareness, can the agent’s person-level attitudes be a sufficiently large part of the explanation of his non-accidentally right action to make it properly attributable to *him*? If such awareness is necessary for the relevant attributability, then it is also necessary for moral worth.

To give this suggestion a fair hearing, it will be useful to consider what some from the Pro-Consciousness Camp say about attributability and the minimal consciousness condition. Concerns about attributability are, in fact, among the principle motivations for many in the Pro-Consciousness Camp (see, for instance, Levy 2014: 87-108). So, let us briefly examine Neil Levy’s minimal consciousness condition and his discussion of consciousness and attributability (this will also give us an opportunity to examine a consciousness condition from the Pro-Consciousness Camp). I will argue that, while Levy helpfully reminds us that an agent’s person-level attitudes must explain why *x* obtains for *x* to be attributable to the agent, he is mistaken in suggesting that “consciousness of the right-making features of one’s action” (whether understood in the ways we have seen from the Anti-Consciousness Camp or in the way we will see from the Pro-Consciousness Camp) is necessary for an agent’s non-accidentally right action to be explained by the agent’s person-level attitudes. Let us turn to Levy’s proposed minimal consciousness condition.

On Levy’s official statement of the “Consciousness Thesis,” “in order to be morally responsible for their actions, agents must be conscious of facts that explain the valence of its moral significance” (2014: 37). While this might look like just another statement of the view earlier attributed

to Arpaly and Sher (namely: moral worth requires awareness of the nonmoral facts in virtue of which one's action is right), Levy's understanding of "conscious" makes the view different.

For Levy, information is conscious when it is "online" (i.e., guiding the agent's behavior) and "personally available"—that is, "when the agent is able to effortlessly and easily retrieve it for use in reasoning" (2014: 33). And, information is "available for easy and effortless recall if it would be recalled given a large range of ordinary cues: no special prompting (like asking a leading question) is required" (2014: 34). So, Levy's minimal consciousness condition relies on the cognitive relation of *reportability*—whether an agent's right action is morally worthy crucially depends on whether the agent can, at the time of acting, *bring to mind* the features of his action which make it right and to which he actually responds, even if the agent does not deliberate upon that information or subject it to occurrent focus. I may perform a rescue that is so cognitively demanding that all my attention is given to mechanical operations ("turn this dial 90 degrees now, flip switches B and C now, ..."). I am not then thinking about the fact(s) in virtue of which my action is right (e.g., the fact that I am rescuing some people, et cetera). However, I can satisfy Levy's minimal consciousness condition (and my action can be morally worthy) because, at the time of performing the rescue, it is true of me that I *could* easily report the facts that guide my behavior and in virtue of which my action is right (see Levy, 2014: 34).

Levy is not without philosophical company in endorsing a reportability requirement on the kind of agency needed for moral worth. John Doris writes that

Where the causes of her cognition or behavior would not be recognized by the actor as reasons for that cognition or behavior, were she aware of these causes at the time of performance, these causes are *defeaters*. Where defeaters obtain, the exercise of agency does not obtain (2014: 64-65).

Similarly, Fischer and Ravizza, in clarifying what it is to act for a reason in the way relevant to moral worth, approvingly reference Robert Audi (1986) and suggest that

In order for an agent to act for a reason, *r*, it is not necessary that the person deliberate and formulate *r* as his reason for acting; roughly speaking, it is enough that he would give *r* as the reason for his action, if he were asked for an explanation (1998: 64).

So, on this view, the minimal consciousness condition is that moral worth requires that the agent easily be able to report the morally relevant facts that guide her behavior and in virtue of which her action has its moral significance. Is this plausible?

It is not. For now familiar reasons, this condition sets overly-sophisticated requirements on moral worth—we cannot reasonably expect ordinary, intuitively morally worthy agents to satisfy this condition. Levy writes as though it is easy for agents, even children like Huck, to satisfy this condition:

[Huck] rejects his [explicit moral] principles altogether, but continues to guide his behavior by reference to the facts upon which moral principles genuinely supervene. *All of this he does consciously*. He lacks only the concepts to perspicuously describe what he does; he lacks nothing in the way of \*awareness of what he does (2011a: 260).<sup>14</sup>

The suggestion, here, is that, were we to interrupt Huck mid-action and ask him what he is doing and why, he would easily be able to report the facts in virtue of which his action is morally right (he just would not use explicitly moral concepts).

But this interpretation of Huck is highly implausible. *The fact that explains the moral rightness of Huck's action* is, as we have seen, some long, unwieldy, conjunctive fact. To satisfy Levy's condition, Huck would need to be capable of bringing to mind what I had previously formalized as “[(1) & (2) & (3) & (4) &...]”. And, surely, we cannot expect ordinary, morally worthy agents to be able to do this. Here is another way to put the point: ethics is hard. If we want there to be ordinary agents whose morally right actions can be morally worthy, it had better not be the case that moral worth requires the ability to bring to mind the set of nonmoral facts that jointly explain an action's moral rightness.<sup>15</sup>

So, though he uses a different cognitive relation in articulating his view, Levy's version of the view that moral worth requires consciousness of the right-making facts is just as problematic as the

---

<sup>14</sup> Note that “\*awareness” just picks out consciousness in Levy's “personal-availability” sense (see Levy 2011a: 247).

<sup>15</sup> *Psychology* is hard, too. We cannot reasonably expect ordinary, intuitively morally worthy agents to be able to bring to mind the full set of morally relevant information to which they respond. Indeed, Doris—who endorses a reportability requirement on moral worth as we see above—argues for agency skepticism precisely on the grounds that we cannot reasonably expect agents to be able to bring to mind the full set of information to which they respond (see 2014: 41-77).

version assessed earlier from the Anti-Consciousness Camp. However, a benefit from evaluating what someone from the *Pro-Consciousness* Camp says is that we are presented with an explicit attempt at motivating the view (philosophers from the *Anti-Consciousness* Camp are more just *inadvertently* committed to a role for consciousness). And, indeed, Levy explicitly suggests that satisfying his purported minimal consciousness condition is required for attributability:

In order for our actions to express our evaluative agency [in the way required for moral worth], *we* must be able to assess the moral significance of our actions for consistency with the beliefs, desires, goals, and commitments (and so on) that together constitute our evaluative agency ... (2014: 107).

When agents are aware neither of the mental states that are responsible for the moral significance of an action, nor of that moral significance itself, ... *the agent* cannot assess either for consistency or conflict with their personal-level beliefs. The action therefore does not express their evaluative agency [in the way that moral worth requires]" (2014: 102).

Levy offers an intuitive picture of attributability. On any plausible view, if *x* is to be attributable to *me*, it must be the case that facts about *me*—about my person-level beliefs, desires, goals, or values—are a sufficiently significant part of the explanation about why *x* obtains. Were it true that consciousness of the facts that make one's action right were the only route through which one's person-level attitudes could play a significant role in the production of one's non-accidental right action, then such consciousness would be necessary for attributability (and, thus, for moral worth). But this is not the only route. If 1) an agent's background desires and values can influence how one accesses and processes basic sensory cues from one's environment (rather than *the facts in virtue of which one's action is right*) and 2) this influence can yield reliably produced right action in various situations, then consciousness of the facts in virtue of which one's act is right is not required for a non-accidentally right act to be properly attributable to the agent. In what follows, I show that (1) and (2) are true.

It is worth pausing to consider where things stand. We have seen that a popular view represented across the *Pro-Consciousness* and *Anti-Consciousness* Camp is that moral worth requires consciousness (in some sense) of the facts in virtue of which one's action is morally right. I have cast

doubt upon this as a necessary condition for moral worth: the condition is excessively restrictive, as satisfying it requires *much* more cognitive and moral sophistication than we can reasonably expect of ordinary, intuitively morally worthy agents.<sup>16</sup> After considering modifications of this consciousness condition that seem less restrictive, we saw that satisfying those condition is at least not *sufficient* for making a right action morally worthy. And, when we saw what can still be missing in cases where someone acts rightly and satisfies these consciousness conditions, the resources we called upon to “close the gap” pointed toward a general explanation of morally worthy actions that had nothing *directly* to do with satisfying these consciousness conditions. We should think that satisfying any of these consciousness conditions is necessary for moral worth *only* if (as a matter of empirical fact) satisfying them is the only way that an agent’s values and desires can explain why a non-accidentally right action was produced. But I have been suggesting that such a view relies on mistaken assumptions about non-accidentality and attributability. Awareness of *something* from one’s environment may be necessary for one’s values to explain why one acted non-accidentally rightly in one’s situation, but attention to how informational access and processing works for creatures like us suggests that such awareness need not be of anything remotely as robust as *the facts in virtue of which one’s action is right*.

Bearing all this in mind, I will now offer an account of moral worth that gives a plausible and necessary role to consciousness, avoids the over-sophistication worries, and secures non-accidentality and attributability without requiring that the agent be conscious in any sense of the facts in virtue of which her action is right.

#### 4. The Value-Secured Reliability Theory of Moral Worth

---

<sup>16</sup> Of course, we could be steadfast in endorsing this as the minimal consciousness condition and just become skeptics about moral worth. To a degree, that is the route taken by Neil Levy (2011b), John Doris (2014), and Michael Zimmerman (1997). But we should take the skeptical road only if it really is true that non-accidentality and attributability require satisfying this deeply demanding condition. And, as I have suggested (and argue in further detail in the next section), neither non-accidentality nor attributability requires the satisfaction of such a consciousness condition.



Here is the proposal.

The core idea is that, since moral worth requires non-accidentality and attributability, what moral worth ultimately involves is a reliable tie to the right that is sufficiently secured by person-level values. I suggest that an agent's action is morally worthy just to the extent that its production is explained by her value-secured reliable tie to the right. Call this the "Value-Secured Reliability Theory" of moral worth.

Obviously, "value-secured reliability" is a technical term. So, what is it, and why should it matter to moral worth?

An agent has a value-secured reliable tie to the right (in a context) just to the extent that

- 1) she is reliably disposed to perform the morally right action in situations like the one under consideration, and
- 2) that reliability is secured or explained by the influence her person-level values have on shaping (perhaps unreflectively) the inputs and outputs of her cognitive and decision-making systems. (That is: *because* of her person-level values, the agent is reliably fed informational inputs from her environment that, when processed, reliably output morally right action).

An agent's *particular* action is explained by her value-secured reliable tie to the right just to the extent that

- 1) her awareness of the information from her environment that served as inputs to her action is explained by the fact that her person-level values have shaped her patterns of awareness to make her reliably aware of information that, when processed alongside her person-level values, reliably yields right actions in such situations; and
- 2) the decisional output of the processing of that information is explained by the fact that, given the agent's person-level values, she will reliably perform right actions when *that* kind of information is processed.

In a moment, I will walk through an example to illustrate the theory and to show how it yields a minimal consciousness condition. But let me first clarify the theory and explain how it is well-equipped to secure non-accidentality and attributability.

Since the account states that an action is morally worthy *to the extent* that it is explained by a value-secured reliable tie to the right, it captures the intuitive idea that moral worth comes in degrees.<sup>17</sup> Without resolving the geometry of moral worth, we can note that *how* morally worthy an action is will be a function of: the *extent* to which the action is explained by the value-secured reliable tie to the right (as opposed to being explained by self-interested motives, or by the nice smell of the baking bread<sup>18</sup>), of the extent of the *reliability*, of the extent to which the reliability is explained by the agent's values, and of the extent to which the reliability is tied to the *right* (as opposed to the more or less nearly right).

The account secures *non-accidentality* because of its reliability condition. It secures *attributability* because the reliability is, itself, secured or explained by the influence of person-level values (as opposed to some God-hand tinkering with the environment). The Value-Secured Reliability Theory thus has a natural way of explaining key features of moral worth.

So, what minimal role is given to consciousness on the Value-Secured Reliability Theory? And what sense of “consciousness” is relevant to the theory? The sense of “conscious” that is integral to achieving value-secured reliability is roughly what Block (1995) has in mind by “access consciousness” and what Baars (1988) has in mind by “global broadcast.” Information is access conscious or broadcast

---

<sup>17</sup> I argued earlier that it is too restrictive to suggest that moral worth requires awareness of *the set of nonmoral facts in virtue of which one's act is morally right*: we cannot reasonably expect ordinary agents to be aware of this full set of nonmoral facts. But, if moral worth comes in degrees, can a fan of the Arpaly-style minimal consciousness condition claim that awareness of some core *subset* of the right-making nonmoral facts is both 1) perfectly achievable and 2) necessary for one's act to be morally worthy to some less than fully maximal degree (but necessary nevertheless for some lower degree)? I argue “no” (it may well be achievable, but it is not *necessary* for moral worth, even of a lower degree). This is because, so long as one agrees with me about the case of the praiseworthy conversationalist introduced toward the end of Section 3 and further developed later in this section, then one agrees that moral worth (even of a low degree) does not require awareness of some subset of the right-making facts. In principle, one's right act can be morally worthy even when one was aware of nothing more robust than quite minimal sensory cues (rather than *some subset of the nonmoral facts in virtue of which one's act is right*). I thank an anonymous referee at *Ergo* for pressing me to extend the decreed understanding of moral worth to a reevaluation of Arpaly's minimal consciousness condition.

<sup>18</sup> See Doris (2014) for a discussion of a range of influences that might undercut person-level explanations of behavior.

in this sense when it was at least momentarily attended to and thereby made widely available for consumption by sub-personal mechanisms in the agent (to a decision-making system, a judgment-making system, et cetera). Once the information is made widely available, it may be processed by mechanisms in the agent in ways that guide her behavior and judgments without the agent's being aware of *how* her behavior and judgments have been so shaped. In what follows, I illustrate the theory and show how person-level values can shape the inputs and outputs of a decision-making system in such a way that makes a right action both non-accidental and attributable to the agent, *even* when the agent is not aware of the right-making facts.

Note, first, that an agent's person-level attitudes can play a significant role in determining what information is and is not made widely available. This is borne out in commonsense observations: the person who cares about cleanliness is more likely to notice the dust than the person who does not care about cleanliness (Arpaly 2003: 83). It is also uncontroversial in discussions of global broadcast theories of consciousness that person-level attitudes can affect what information is broadcast: such top-down influences on patterns of awareness occur "when one's goals or interests direct attention to one aspect of the stream of current sensory processing rather than another" (Carruthers 2011: 48). Here is, thus, already one meaningful interaction between an agent's person-level attitudes and the information from her environment of which she is aware.

Now, when an agent's person-level attitudes play this role in determining what information is and is not momentarily attended to and thereby made widely available, the fact that one's person-level attitudes play this role need not itself be widely available. This point should be uncontroversial. A good conversationalist's background values (that, say, people feel heard or are not made to feel unnecessarily uncomfortable) may direct her attention to various features of her environment (subtle facial cues, et cetera) even while she is altogether unaware of the role that her goals play in guiding her attention.

Once some information is attended to and made widely available, that information can be consumed by a decision-making system that has direct access to some of the agent's person-level attitudes (her beliefs and goals) without those person-level attitudes *themselves* needing to be attended to. As Carruthers notes, "we should expect ... decision-making systems to be capable of accessing some of the subject's beliefs and goals directly, without the latter needing to be reached through global broadcast" (2011: 53). Were this *not* true, an agent would need to attend to her relevant standing beliefs and goals any time she performed some action that drew upon them—agents would be mentally exhausted by the end of breakfast.

So, the decision-making system can output a decision by processing the access conscious information from her environment and the agent's person-level attitudes together. Insofar as the decision-making system outputs a decision by processing both the access conscious information and the agent's person-level attitudes together, the agent's person-level attitudes play an important role in determining how the agent responds to information. To use a simple example: the decision-making system may draw directly on my goal to warm up, so that, when imagistic content pertaining to the mug of green tea and the glass of ice water on the table are briefly attended to (when I see both on the table), the decision-making system processes all of this and outputs a decision to grab the mug of tea rather than the glass of water.

Of course, *that* and *how* the decision-making system has processed some access conscious information alongside some person-level attitudes is not necessarily, itself, access conscious. In the example above where I grab the mug of tea rather than the glass of water, I need not be aware of or have attended to the fact that my goal of warming up was processed alongside sensory information pertaining to the drinks on the table. Now, in this case, I could probably tell you correctly after the fact why I grabbed the tea instead of the water: "I wanted to warm up." But, importantly, not all outputs of the decision-making system are like that. A good conversationalist—a good listener and

conversation partner—pauses at the appropriate moments, changes subjects at the appropriate moments, interjects at the appropriate moments, and so on. Moreover, some of the best listeners/conversationalists are unaware of the extent to which their background values (that, say, people feel heard or are not made to feel unnecessarily uncomfortable) shape their conversational patterns. One might correctly tell Roscoe after seeing him navigate several especially emotionally loaded and complicated conversations: “wow, Roscoe—the way you do x, y, and z in conversation is perfect—I’m going to start trying to do x, y, and z.” And Roscoe might sincerely respond: “huh, are x, y, and z things that I do?”.

Let me continue with the example of Roscoe the Conversationalist to illustrate the joint significance of the considerations adduced thus far. Roscoe’s person-level values may reliably make widely available relevant sensory cues: content pertaining to subtle facial expressions and speech inflections reliably correlated with an interlocutor’s feeling of discomfort are reliably made available to mechanisms in Roscoe in part because of his background person-level values (e.g., his desire that people are not made to feel unnecessarily uncomfortable). Moreover, when sensory cues like *those* are made widely available to Roscoe, a right action—to, say, wait a beat in the conversation—is reliably produced from the unconscious processing of those sensory cues alongside his person-level values.

When Roscoe waits a beat in conversation, his action is morally worthy: it is explained by a value-secured reliable tie to the right. It is no accident that he waits a beat at the appropriate moments, and this non-accidentality is explained to a large degree by the influence that his person-level values unreflectively have on shaping his access to and processing of information.

In this test case, Roscoe need not be conscious (in any sense) of the fact that waiting a beat right now (or at a moment like this) is right. He may not even be conscious (in any sense) *that* he waits a beat, and he may have no nuanced, readily articulable beliefs about how conversations should go.

Moreover, Roscoe need not be conscious (in any sense) of the fact(s) in virtue of which it is morally right to wait a beat in the conversation. He may not be conscious (in any sense) of the fact that (say) Nina wishes to add a qualification to her most recent point. The wide availability of various sensory cues (furrowed brows, voice inflections, et cetera) can, itself, reliably lead to the outputting of a morally right decision (to wait a beat) given Roscoe's background values. Access consciousness plays a necessary role in enabling Roscoe's person-level values to guide his response to his environment, but access consciousness of *the fact(s) in virtue of which his action is right* need not have played any role.

On the Value-Secured Reliability Theory, an agent need only be aware of (or, have momentarily attended to and thereby have widely available) *whatever* minimal information from her environment will, in virtue of being processed against her particular set of background values, enable her to reliably perform right actions. What *specific* information an agent must have widely available if her action is to be explained by a value-secured reliable tie to the right may vary depending upon her values, the strength of those values, and how morally complicated the action is. As we see in the case of Roscoe, this information may sometimes involve nothing more robust than basic sensory cues (ones about which the agent may be in no position to say anything concerning how they guided his behavior). But access consciousness of some information from one's environment *is* necessary for moral worth on this account—for, even when an agent's right act is reliably produced, it is only when its production has been shaped by the agent's person-level attitudes that the act can be *attributable* to the agent. It is only when some information from one's environment is made widely available to mechanisms in the agent that the agent's person-level attitudes have an opportunity to be brought to bear on shaping his response to his environment. This does not set the bar too high for moral worth. Depending upon an agent's background values, an agent could act out of a habit and be morally worthy. But this will be so only if 1) the habit is reliably tied to the right in the situation and 2) when the agent acts from the habit, his background values play a sufficiently large role in explaining why

certain information was (rapidly, unthinkingly) made widely available and processed in such a way that generated a right action.

In Huck's case, it is perfectly well imaginable that his decision to keep helping Jim is explained by a value-secured reliable tie to the right. *Because* of his person-level values, in situations like this one, Huck is reliably made access conscious of information that, *when* access conscious, reliably leads to right action. When Huck's action is explained by his value-secured reliable tie to the right, his action is both non-accidentally right and attributable to him. Access consciousness plays a necessary and significant role in this, but none of this requires that Huck be aware of the set of facts in virtue of which his action is right. The Value-Secured Reliability Theory offers a successful account of moral worth that gives a motivated and delineated role to consciousness.

To further spell out the implications of the theory, it will be useful to respond to a potential objection. Consider Kant's famous self-interested grocer. The grocer's business decisions are explained by his desire to maximize profits—it just so happens (we stipulate) that he will maximize profits if he regularly makes business decisions that are morally right. It is intuitive that the grocer's decisions are not morally worthy. This might be a problem for the Value-Secured Reliability Theory. After all, the grocer reliably makes business decisions that are morally right, *and* he reliably makes the decisions he makes because of a person-level value (the value he places on making as much money as possible). He thus appears to have a value-secured reliable tie to the right. Is the theory committed to saying that his business decisions are morally worthy?

It is not. Let me sketch a tempting but ultimately unpromising response first, and then I will offer the correct response.

It is tempting to argue that the grocer's decisions are *not* reliably tied to the right. Presumably, there are nearby possible worlds where his business interests do not align with morality, and in such worlds, the grocer would *not* make the morally right decisions. Moreover, in situations outside of

business, he is perhaps unlikely to treat people fairly. The Value-Secured Reliability Theory would then yield the intuitively correct verdict that his actual business decisions are not morally worthy, since he does not have a value-secured reliable tie to the right.

But, Kant's grocer case is interesting because his behavior *does* seem to be reliably tied to the right (at least in some domain). Any theory of moral worth should allow that an agent can perform genuinely morally worthy actions in *some* domains while having moral blind spots that prevent him from performing morally worthy actions in other domains (imagine the professor who performs genuinely morally worthy actions within the department but is callous with family). We do not want the result that an action is morally worthy only if the agent is reliably tied to the right across *all* domains. While there is certainly an important issue concerning how to demarcate the relevant domains, there is not the space to settle the issue here. So it would be good to have a different response to the case.

A different response draws on the role of *explanation* in the Value-Secured Reliability Theory. Recall that, on the theory, it is not enough that an agent simply *have* a value-secured reliable tie to the right. A particular action is morally worthy just to the extent that its production is *explained* by that value-secured reliable tie to the right. The grocer may well have a value-secured reliable tie to the right. But, when he gives the eight-year-old correct change on Friday as morality requires, this transaction is not *explained* by his value-secured reliable tie to the right. There is a competing and better explanation that undermines this one: namely, the transaction is explained by a value-secured reliable tie to *maximal profits*. An explanation of the grocer's transaction that references a reliable tie to the right is outstripped by (rather than supported or amplified by) an explanation that references a reliable tie to maximal profits. The moral explanation is undercut, as the tie to maximal profits is not plausibly "a part or a symptom" of the tie to the right (see Sturgeon 1992: 100).

Contrast the case of the self-interested grocer with Roscoe, our morally worthy conversationalist. Both characters have value-secured reliable ties to the right. But *Roscoe's* particular



right action (to wait a beat in conversation) is best explained by his value-secured reliable tie to the right. Roscoe's values have shaped his patterns of informational access and processing to reliably produce right actions in situations like the one he is in—that is what most fully explains why he came to have access to the information that served as inputs to his decision and why the processing of that information produced a right action. We could explain Roscoe's action with nonmoral language. We could say that his action is explained by a reliable tie to actions that promote his relevant values (his values that people feel heard, or are not made to feel unnecessarily uncomfortable, or the value he places on the persons around him). But this nonmoral explanation does not undermine the moral one—it *supports* it (see, again, Sturgeon 1992). The tie to the promotion of such values (insofar as these values are for things that matter morally) *is* plausibly a part or symptom of a tie to the right.

The Value-Secured Reliability Theory is *not* committed to saying that wherever there is a reliable tie to the right, there is moral worth. It shows, instead, that an action is morally worthy just to the extent that it is *explained* by the agent's value-secured reliable tie to the right.

## 5. Conclusion

The Value-Secured Reliability Theory offers a plausible account of moral worth and shows, *contra* the popular view from the Pro-Consciousness and Anti-Consciousness Camps, that moral worth does not require consciousness (in any sense) of the facts in virtue of which one's act is right. More than this, it offers a rich theoretical framework that may prove useful for thinking about a range of philosophically interesting phenomena that involve some success that is non-accidental and attributable to the agent.<sup>19</sup> In the case of moral worth, non-accidentality and attributability are secured

---

<sup>19</sup> Epistemologists may care about when it is attributable to *me* that my belief was reliably produced (see Sosa 2007, 2015). The philosopher of art may care about when it is attributable to *me* that my painting is a good piece of art (see Wolf 2015). Though it is beyond the scope of this paper to extend the Value-Secured Reliability Theory to these domains, we might speculate about how this would be done. Because of the influence that an agent's person-level values have on shaping her patterns of informational access and processing, an agent might be reliably tied to the right, or to making good artistic decisions, or to forming beliefs through truth-reliable processes. When an agent's right action (or artistic decision, or

when, *because* of the agent's value-secured reliable tie to the right, certain (perhaps mere sensory) information from her environment is made widely available and then processed to produce a right action. Consciousness is still necessary for moral worth because it is the vehicle through which an agent's value-secured reliable tie to the right is brought to bear upon her response to her environment. As we saw with Roscoe the Conversationalist, an agent's right action can sometimes be explained by a value-secured reliable tie to the right (and, thus, be morally worthy) even when the agent is conscious of neither the fact that he acts rightly nor of the fact(s) in virtue of which his action is right, and even when he is incapable of explaining how his values might have shaped the production of his right action. The role of consciousness in a theory of moral worth is no bigger nor smaller than the role it plays in making our actions explained by a value-secured reliable tie to the right.

### **Acknowledgments**

I wish to thank Sophie Horowitz, Hilary Kornblith, and Katia Vavova for numerous conversations about the material in this paper and for extensive comments on earlier drafts. I also thank two anonymous referees at *Ergo* for helpful questions and suggestions. Finally, I am grateful to the audience in attendance for my presentation of an earlier version of this paper at the University of Pavia.

### **References**

Arpaly, Nomy. 2003. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press.

———. 2015a. "Huckleberry Finn Revisited: Inverse Akrasia and Moral Ignorance." In

Randolph Clarke, Michael McKenna, & Angela M. Smith (ed.s), *The Nature of Moral Responsibility*. Oxford University Press.

---

reliably produced belief) is explained by such a value-secured reliable tie to the right (or to the artistic good, or to the truth-reliable), there may be good reason to believe that the success is both non-accidental and attributable to the agent.

- . 2015b. “Consciousness and Moral Responsibility, by Levy, Neil.” *Australasian Journal of Philosophy* 93 (4): 829-831.
- Arpaly, N. and Schroeder, T. 1999. “Praise, Blame, and the Whole Self.” *Philosophical Studies* 93 (2):161-188.
- . 2013. *In Praise of Desire*. Oxford University Press.
- Audi, Robert. 1986. “Acting for Reasons.” *Philosophical Review* 95 (4): 511-546.
- Baars, Bernard J. 1988. *A Cognitive Theory of Consciousness*. Cambridge University Press.
- Bennett, Jonathan. 1974. “The Conscience of Huckleberry Finn.” *Philosophy* 49 (188):123-134.
- Block, Ned. 1995. “On a Confusion about a Function of Consciousness.” *Brain and Behavioral Sciences* 18 (2):227-247.
- Carruthers, Peter. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford University Press.
- Doris, John M. 2015. *Talking to Our Selves: Reflection, Ignorance, and Agency*. Oxford University Press.
- Fischer, J.M. and Ravizza, M. 1998. *Responsibility and Control*. Cambridge University Press.
- Levy, Neil. 2011a. “Expressing Who We Are: Moral Responsibility and Awareness of our Reasons for Action.” *Analytic Philosophy* 52 (4):243-261.
- . 2011b. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford University Press.
- . 2014. *Consciousness and Moral Responsibility*. Oxford University Press.
- Sher, George. 2009. *Who Knew? Responsibility Without Awareness*. Oxford University Press.
- Sliwa, Paulina. 2016. “Moral Worth and Moral Knowledge.” *Philosophy and Phenomenological Research* 93 (2): 393-418.
- Sosa, Ernest. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume I*. Oxford University Press.

Press.

Sosa, Ernest. 2015. *Judgment and Agency*. Oxford University Press.

Sturgeon, Nicholas. 1992. "Nonmoral Explanations." *Philosophical Perspectives* 6: 97-117.

Wolf, Susan. 1980. "Asymmetrical Freedom." *Journal of Philosophy* 77 (3): 151-166.

———. 1987. "Sanity and the Metaphysics of Responsibility." In Ferdinand David Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge University Press.

———. 2015. "Responsibility, Moral and Otherwise." *Inquiry: An Interdisciplinary Journal of Philosophy* 58 (2): 127-142.

Zimmerman, Michael J. 1997. "Moral Responsibility and Ignorance." *Ethics* 107 (3): 410-426.