



---

Honors Theses at the University of Iowa


---

Fall 2018

## From Akrasia to Quality of Will: A Critical Examination of Accounts of Moral Responsibility and Ignorance

John Kirchner

Follow this and additional works at: [https://ir.uiowa.edu/honors\\_theses](https://ir.uiowa.edu/honors_theses)

 Part of the [Epistemology Commons](#), and the [Ethics and Political Philosophy Commons](#)

---

This honors thesis is available at Iowa Research Online: [https://ir.uiowa.edu/honors\\_theses/258](https://ir.uiowa.edu/honors_theses/258)

---

FROM AKRASIA TO QUALITY OF WILL: A CRITICAL EXAMINATION OF ACCOUNTS OF MORAL  
RESPONSIBILITY AND IGNORANCE

by

John Kirchner

A thesis submitted in partial fulfillment of the requirements  
for graduation with Honors in the Philosophy

---

Ali Hasan  
Thesis Mentor

Fall 2018

All requirements for graduation with Honors in the  
Philosophy have been completed.

---

Ali Hasan  
Philosophy Honors Advisor

## From Akrasia to Quality of Will: A Critical Examination of Accounts of Moral

### Responsibility and Ignorance.

What is it to be morally responsible? As opposed to being held responsible within other contexts (legal, causal, etc.)? What is unique about moral responsibility *specifically*? Moral responsibility has a distinct *kind* of normative weight; it also requires some sort of *not-so-straightforward* connections between agents and their actions (yet further, between agents and specific psychological phenomena). Moral responsibility is also (at least) a precondition for further moral evaluations of agents, most prominently in determining their praiseworthiness and/or blameworthiness. But beyond noting these skeletal, theoretical parameters, it seems hard to say more about moral responsibility with much confidence.

There is an undeniable opaqueness to these types of moral evaluation and it's inherently difficult to nail down definitive conditions which identify grounds for agents' culpability. However, that being acknowledged, one particularly interesting means of testing conditions for culpability (and more broadly for understanding the epistemic and cognitive aspects to moral evaluations of agents) is to look at and analyze cases of *morally consequential ignorance* (i.e. cases concerning agent's ignorance of their actual wrong-doings). Indeed, I will be focusing heavily on cases with this very content and in doing so it will (hopefully) help enumerate and distinguish the many (already existent) attempts to understand and classify moral responsibility, blame/praise, and culpability.

This work will explore various instances of moral ignorance and will critically examine the viability of certain posited accounts of moral responsibility for actions done under ignorance.

From leading *Akratics*, to various *Quality-of-Will* views (and some accounts in between), there will be an attempt to make sense of and to challenge these distinct views. Seeing wherein one succeeds and another fails will ultimately help to observe something essential (something inter-theoretical) about agents, actions, and how they properly intertwine in an understanding of moral responsibility and in moral evaluations more broadly (e.g. praising, blaming, and exculpating). The aim herein will be mostly critical and problematic commitments will be demonstrated for all the views explored. Moving from exposition on and presentation of the basic conflicts arising in cases of morally consequential ignorance in general, there will be a step-wise examination of Akratic, Quality-of-will, and intermediary accounts. Each account will be shown to have its unique set of problems and strengths. But ultimately, what emerges from this brief and focused survey is a broad and cross-theoretical understanding of the true challenges (two in particular, as will be demonstrated) facing positive accounts of moral responsibility for ignorance.

The specific cases of interest herein, are (as mentioned) those in which agents act in ignorance; especially when it concerns ignorance of some wrong-making feature of their actions, or psychological states. When acting ignorantly, agents still (of course) evoke moral consequences, which can gravely affect the lives of others. But the ignorance behind an agent's actions have a *muddying* effect; so that often it becomes obscure as to *how, if, and when* the actor can be properly connected to the moral consequences of his/her actions and as to how to delineate the various normative demands which might otherwise be acting upon him/her. Such obscurity raises concerns like: could agents (or *ought* they) have avoided a particular outcome?

Are certain actions indicative of agent's genuine moral commitments, or character dispositions?

Also, what *types* of ignorance are exculpatory (if there are any)? And under what

circumstances, and given what standards, do we expect agents to have done otherwise?

Well, to begin to answer these questions we will need to say something about the nature of moral evaluation broadly and we will need to speak to how agents relate to their actions (as fundamentally distinct categories) in moral evaluation.

### **1. Agents, Actions, their Interconnection and Preconditions to Moral**

#### **Evaluation:**

So as to try and reach a minimal point of clarity on moral evaluation in general, the first step will be to understand *how actions and agents differ in moral evaluation*. To begin, it needs be stipulated that: acts (and omissions) are that which can be evaluated as morally wrong, or morally right, but it's agents that can be deemed blameless, blameworthy, or praiseworthy for performing the given act. And to be clear, even very distinctly different accounts want to maintain this same delineation in their frameworks of moral evaluation. So this is to say that, the basis of the disparity between the accounts we'll explore is something else, something more along the lines of, as Zimmermann explores, "Under what circumstances *is* someone praiseworthy for something morally commendable or blameworthy for something morally objectionable?" (227); an account's specific *circumstantial stipulations* are primarily what we're interested in looking at, since this is more precisely the basis of divergence between any account of moral responsibility (not to say that they all don't diverge in various other ways).

A basic shared goal then, of any account of moral responsibility, is to establish precisely how and under what conditions, agents are accountable for their actions (and further yet, how we are entitled to respond to them, should the proper connection between the agent and action be established). That being said, to demonstrate how this evaluative distinction plays out and why it is necessary and so fundamental (again borrowing again from Zimmermann's exploration of this connection [224-228]), we'll start by looking at the absurdity of collapsing the two distinct loci of moral evaluation.

To use (rather loosely) a derivation of Zimmermann's example (see 224); we can all acknowledge that the Holocaust was undoubtedly a terrible act. But *that it was a morally terrible action* tells us nothing more of morally evaluative import. For instance, tornados (which wipe out entire city blocks) are also morally "bad", at least owing to their morally bad consequences, but a weather event and the Holocaust are very obviously different in important ways (at least when it comes to moral evaluation). Why? Precisely because agents are intimately connected to bringing about (and perpetuating) the Holocaust's horrors, while no one need be (nor can be) held responsible for a weather event's horrors (putting aside instances of wrongly informed meteorologist giving knowingly false forecast, or poorly trained first responders harming storm victims, etc.). That is to say, the tornado's existence is not agent dependent in the same way as the holocaust is.

Thereby, we can rightly and easily identify that the Holocaust is a morally wrong action (or perhaps series of actions), for which some agents ought to be held accountable (unlike tornados). But what if we were to simply choose to hold *you* responsible for it on a whim? I'd

imagine you would not see it as fitting that we do so. But why? Well, because you must, in some sort of circumstantially relevant way, be appropriately tied to the action for it to be fitting for us to claim you are responsible for it. As far as the Holocaust goes, you likely were not even born at the time it occurred; but even if you had been living in Germany at the very same time it had, we still would clearly need some kind of account about how your conduct is connected to this morally wrong act. We (firstly) would have to establish this causally, but more importantly, we then would need to explain: *why you personally are responsible for engaging in the conduct as you did.*

This connection is fleshed out in innumerable ways, given the context of different accounts (as will be seen), but without some type of qualification about *how the action connects to the actor*, attributions of responsibility cannot properly take place. We clearly cannot on a whim decide that we resent you, or admire you, or hold you accountable, or excuse you, because there is some bad consequence that merely causally happens to be connected to you.

However, perhaps that is enough said abstractly on the matter of morally evaluative agent/action interconnection for the time being. To truly understand the problematic nature of the kind of moral evaluation at hand, we need some case of *morally consequential ignorance* to sink our teeth into, which will more concretely enumerate the relevant distinctions at play and which will help us more productively see the most pertinent concerns for disambiguating the concepts at play in moral responsibility.

## **2. Holly Smith and a Case of Morally Consequential Ignorance:**

The initial case I will be presenting, so as to accomplish a rudimentary and more tangible explication of the types of morally evaluative ignorance we will be exploring throughout this work, is a case that already exists in numerous iterations within the literature on moral responsibility for ignorance (a subset of moral theory often alternatively called: *culpable ignorance* [See the works of Gideon Rosen and Holly Smith for the precedent of the terminology]). The following case is, more specifically, a derivative of Smith's initial case from her highly influential work entitled, "*Culpable Ignorance*".

**a. The Case of Doctor S.**

General practitioner, Doctor S., is seeing an expectant mother (Ms. P) for complaints of excessive morning sickness. Doc. S. accordingly administers to his patient a particular intravenous drug (*Drug T.*) to help alleviate her symptoms and to lessen the severity and frequency of her morning sickness. However, unbeknownst to Doc. S, there was a recent medical case-study published, which strongly implicated Drug T. as having consistently caused birth defects, when administered to pregnant mothers. Doc. S had not read the recent medical literature on Drug T. Also, Doc. S had at his disposal a wide selection of alternative medications (none of which had been implicated as dangerous), which would have been equally effective for treating the very same presenting symptoms of Ms. P; but regardless, Doc. S administered the one and only implicated drug from those available within in the clinic. Unfortunately, as a result, Ms. P's child is born malformed and the cause (let us say indisputably and solely) is shown to be the pre-natal presence of Drug T., which Doc. S chose to administer.

**b. An Initial Response:**



So what does one make of Doc. S and his role in leading to malformation of Ms. P's child? Initially, it's clear enough that Doc. S is *causally responsible* for bringing about the malformation though he had no idea at the time what consequences his actions/choices would bring about. Secondly, it was *possible* (indeed easy) for Doc. S to do otherwise than he had. He had other medications available to him at the time he decided to treat Ms. P, which would have treated the same symptoms and condition effectively. Alternatively, he could have not treated her at all and refrained from administering any drug, or course of treatment, whatsoever.

What's important to note though is that *Doc. S did have a chance to do otherwise*, and his harming Mrs. P. and/or her child was not *an inevitability*. Doc. S could also have read the recent literature concerning Drug T. prior to administering it to Ms. P. That is, separate from the action of administering the drug itself, there existed *a previous action* that he could have performed so as to have avoided the same ghastly results. That previous action is where Smith hones in, and indeed, a longstanding camp in the literature traces one's responsibility for such actions, if any, to responsibility for one's ignorance.

### c. **Benighting Actions:**

Holly Smith provides us with an applicable and ready-made terminology to help properly delineate and identify the separate features at play in cases such as these: “. . . the relevant cases [including ours] all involve a sequence of acts: an initial act, in which the agent fails to improve (or positively impairs) his cognitive position, and a subsequent act in which he does wrong because of his resulting ignorance. [. . .] Call the initial act the ‘benighting act’ and the subsequent act the ‘unwitting wrongful act’” (547). So the “*benighting act*” would be not

reading the relevant journal entry on Drug T., which “*impaired*” the “*cognitive position*” of Doc. S when eventually examining a pregnant patient; then, the “*unwitting wrongful act*” would be treating said patient (Ms. P) with a harmful drug (harming her or her child as a result).

**d. Zimmerman, Rosen, and Culpability for Akratic Acts:**

If we want to hold Doc. S accountable for the role he played, we must focus on the nature of the *benighting action*, since it’s apparent from the outset that it isn’t the action done under ignorance --administering the drug-- that strikes us as inherently irresponsible. What seems to (instead) make Doc. S culpable for his conduct, is particularly *his failure to keep appropriately abreast of the literature concerning the drugs he might/did administer*. Is it then, that we think Doc. S has failed some normative standard, regarding his epistemic practices, which it was reasonable for us to have expected him to meet?

Now, to be clear, before we go any further into the case-analysis, had Doc. S read the relevant case-study on Drug. T before examining Ms. P and chose to administer it anyway, such conduct *would be non-ambiguously culpable* and would exist on an entirely differently level. Acting voluntarily and with foresight of morally bad and probable consequences, so as to *directly* (and with intention) do another harm, is plainly a most base kind of moral conduct. But still, even though Doc. S didn’t know the very palpable risk of harming his patient with the drug as he administered it, there seems to be a moral failure nonetheless. *This* moral failure is what we need to further examine.

Perhaps the nature of the infraction is, as it is often referred to in the literature, its being a case of *akrasia*. But what is *akrasia* and why would it be relevant for moral evaluations of ignorance? Well, *akrasia* in action (to borrow Nomy Arpaly and Timothy Schroeder's concise definition) is: ". . . to believe one course of action to be, all things considered, the right one, yet to take some other course of action instead." (162). And Doc. S, conceivably, could be acting akratically in his benighting action (i.e. in not reading the case-study on Drug T). Doc S. would have been acting akratically if we assume that: **1)** he had known that doctors, such as he, ought to be minimally diligent in regards to their medical practice and the drugs they use, and **2)** he was aware that he was, at the time of the benighting action, falling to act in accordance with this knowledge (thereby knowingly risking a morally weighty consequence if he were rendered ignorant of the proper medical knowledge at some later time), and finally **3)** that he still chose to act otherwise than to properly inform himself. It seems if these conditions were met, that we should hold Doc. S fully responsible for the harm he caused, because he acted in full knowledge of the wrongness he might later cause.

**e. Zimmerman's Conditions:**

But before we were to say that Doc. S might be morally responsible, we would need to be clear on the theoretical account of moral responsibility we would be committing to (in evaluating him by an akratic standard). To give one of the more concise standards for culpable akratic action, we can look to the contributions of Michael J. Zimmerman. For instance, he provides us with two "*individually necessary and jointly sufficient*" conditions (which he claims are derived from pre-existing Aristotelian literature on the matter [219]) that identify the

proper parameters of culpable action (219). Zimmerman specifies accordingly, that: “First, there is a freedom condition. Was the person in *control* of that for which he or she is allegedly responsible? Second, there is an epistemic condition. Did the person *know* what he or she was doing?” (219).

Doc S. was clearly in control of his actions (for both the unwitting-wrongful and benighting actions, though perhaps an action done ignorantly, i.e. unwittingly, is constrained in a sense). However, the epistemic condition is only (unambiguously) met in the benighting action. Doc S. did not know that certain morally significant consequences would ensue from using the particular drug he chose to treat Ms. P with. He lacked knowledge of the wrongness of his act. Indeed, it is called an “unwitting” act for a reason. Abiding by these conditions, only Doc.’s benighting action meets them both. So if we’re to pin down the locus of moral evaluation appropriately, it would be the benighting action where we could say something about Doc. S. and the harm he caused. Zimmerman’s conditions also explain that we need to ensure that circumstantially an agent’s *sufficient awareness* and *sufficient control* were met in holding him directly culpable for some action (this of course is the type of connection/pre-condition we spoke of in the first section).

However, it’s important to be clear that the akratic conditions just mentioned entail only one’s *direct* culpability (in being met that is). One can also be *indirectly culpable* for (potentially a host of) other actions, if at some preceding point one was *directly culpable*, i.e. meeting both of Zimmerman’s conditions, and if the action that followed was a result of the preceding act. If (to use the example at hand) we were to show that Doc. S had acted akratically in skipping over

the recent literature on the drugs he uses in his practice, we would say both that: 1) he is *indirectly culpable* for the *unwitting wrong* of harming Mrs. P., and 2) he is *directly culpable* for the *benighting action* of skipping over the literature he knew he should have read before seeing his patients (see Randolph Clarke, 225 for a cogent discussion of the direct/indirect distinction).

To summarize, should we want to hold Doc. S responsible, we now see it must be only (non-derivatively so) for his *benighting action*, because it *alone* meets the two akratic conditions. However, some reject or modify the conditions as specified (for instance, Clarke has his own two conditions, but modifies the 2<sup>nd</sup> [awareness] condition to being a broadly “cognitive” condition [241]). Similarly, Gideon Rosen (though certainly an akratic through-and-through) provides an account which more heavily leans into the epistemic condition of moral responsibility and which modifies it by enumerating corresponding duties in relation to it, which further nail down the class of morally evaluative and culpable actions.

**f. Rosen and “Epistemic Obligation”:**

In the evaluation of Doc. S above, we contended that he would be potentially morally responsible for the harm he brought about, if we could say that he acted akratically in his benighting act, which preceded the act of administering his inadvertently harmful treatment. The essential idea (again) was that Doc. S was importantly ignorant of his wrong-doing, and while persons should otherwise be excused for acting wrongly from ignorance, we might say that if his benighting action was truly akratic, then Doc. S (in some sense) *inappropriately maintained* his beliefs about how best to treat his patients (i.e. because he was inadequately informed and not up to date on the medical literature concerning his domain of practice).

Gideon Rosen provides a helpful observation, in this vein, concerning what he calls our “*epistemic obligations*” (63), which expand upon the Zimmermannian awareness condition above by presenting a set of morally obligatory duties in relation to maintaining proper awareness. Rosen notes accordingly:

“We are under an array of standing obligations to inform ourselves about matters relevant to the moral permissibility of our conduct: to look around, to reflect, and to seek advice, and so on. [. . .], whenever a person acts badly from ignorance the question will always arise whether [one] has discharged these epistemic obligations.” (63).

So thereby, in tracking why akratic action itself might be blameworthy, it may be an explanation along the line of Rosen’s principle above. To put the moral evaluation of akratic action in more specific context now, perhaps Doc. S’s akratic benighting action would be particularly culpable *because he did not “discharge” his “epistemic obligations”* as a practitioner of medicine (practitioners of medicine being naturally under more stringent than normal obligations to inform themselves, in virtue of their profession).

Of course, it is important to note conversely, that if Doc. S had (somehow) been very poorly trained about what epistemic obligations he had as a doctor and so he did not realize that he ought to inform himself in the morally relevant particular ways (e.g. by keeping up on the recent literature on the drugs he administers), then we would not hold him culpable. For, by his own lights, he did what he thought he was obligated to do. Though we can acknowledge intersubjectively that someone might normally have such-and-such an obligation to inform themselves, if an agent was non-culpably ignorant of the fact that they ought to have been

aware given (X) circumstance, we still can't hold them culpable. For as Rosen notes, "Whenever an agent acts from ignorance . . . he is culpable for the act only if he is culpable for the ignorance." (64).

### **3. In Opposition to the Akratics: Quality-of-Will Theory, Revision, and Finn:**

While the akratic account of moral responsibility seems (*prima facie*) to provide a fully sufficient account of culpable ignorance and even blame in the case of Doc. S., there are other cases of morally consequential ignorance which aren't so easy to evaluate with merely an Akratic framework for understating moral responsibility and praise-/blameworthiness. Arpaly and Schroeder's case of Huckleberry Finn (or for brevity's sake, the case of *Finn*) is (at least) one such instance. Namely (to give the oversimplified problem), the *praise* we intuitively want to give the agent in the following case (i.e. Finn) comes about, not in spite of, but specifically because of his akrasia. He acts freely and in contradiction to what he believes to be right; but nonetheless, it's hard to deny that Finn not only did that which is obligatory, but (further yet) that which is *praiseworthy*. So if acts of akrasia can be fully *praiseworthy* (as it will seem in the following case), then something about moral evaluation is being missed by account which embed culpability within it.

A more detailed analysis of how Finn presents a problem for the akratic account (and which consequences are entailed in an account with such an orientation in moral evaluation) will follow, but it helps (hopefully) to hone in on the basis of the problem for Akratics before looking at the case of Finn itself.

**a.) Revision:**

In addition to the problems that come about in case of Finn, another common objection raised in opposition to akratic accounts of moral responsibility is that they stand to (perhaps too radically) revise our commonplace moral intuitions about when we can hold persons culpable. Namely, those in opposition claim the Akratic Account leads to *far too few* instances of genuinely culpable ignorance. Randolph Clarke notes according to this contention, that: “. . . the main problem posed by the argument for revision [i.e. the Akratic Account] concerns the scarcity of bases for blameworthiness for unwitting wrongful conduct. It is not that often, when one engages in blameworthy conduct, that one is aware of the wrongness of what one is doing. For there is little clear-eyed wrongdoing.” (238). Indeed, in Rosen’s seminal work “Culpability and Ignorance”, it’s worth noting that (through an akratic account of culpability) he ends up exculpating all kinds of commonly blameworthy persons, e.g. from an ancient slave owner to a sexist (see Rosen, 64-69).

However, though it’s certainly important to recognize this objection, not much more countenance will be given to it in our examination herein. For although preserving commonplace moral intuitions is (other things being equal) a dialectically valuable thing for any account to do, this objection has decidedly less force than the case of Finn. So our prioritization in exploring the case of Finn in greater depth (while admittedly only briefly mentioning this other objection) is not to off-handedly dismiss it, but rather to motivate more completely what I hope to prove is the more substantive and dialectically interesting objection to the Akratics.

**b.) Finn and Akrasia:**



The case of Finn is probably roughly familiar for most, as it's one derived explicitly from Mark Twain's famous character and story. Nonetheless, certain aspects of Finn's portrayal are more morally significant than others.

Finn befriends an escaped slave (named Jim); however, ". . . his conventional southern moral convictions tell him clearly that he should proceed to return the slave [Jim] to his lawful owner." (Arpaly and Schroeder 162). Yet, despite his desire to act morally and his conviction that he *morally ought (all-things-considered)* to return Jim, he simply can't bring himself to do so. He thinks he is weak for failing to do as he sees morally fit and bemoans his immorality.

#### **4. A Reorientation in Culpability:**

So perhaps it seems that those opposing the Akratists get something importantly right about the Huckleberry Finn case. For it is precisely the fact that Finn cannot bring himself to do what he thinks to be all-things-considered morally right (i.e. akrasia pure and simple), that we want to excuse and even praise him. As Arpaly and Schroeder note of the basic intuition at play in Finn's case, "The gut feeling is that an inverse akratic like Huckleberry is praised for his 'heart', i.e. for the very desires that motivated his akratic action." (168). Thereby Quality-of-Will theorist's (like Arpaly and Schroeder) undermine the akratic account by shifting the locus of moral evaluation away from merely finding a *voluntary and knowingly wrong action to source one's blame/praise by*, moving instead to aspects of the actor's "*motivations*" or *character*.

The (as it is called) "*Whole Self*" account from Arpaly and Schroeder (and their case of Finn) suggest that it is perhaps psychological states, or cognitive features (e.g. "heart", desires,

and/or motivations) which *are* actually morally evaluative; in fact this particular contention is at the heart (not to mix terms) of *any and all Quality-of-Will* accounts more broadly. Furthermore, for said accounts it's specifically that an action/omission is (in and of itself) taken to have some *implicit content*, which is manifested, or connected to, some type of morally evaluative cognitive feature (as is seen in the case of Finn). This change in focus (i.e. from attempting to trace the *unwitting-ness* of an apparent *unwitting wrong*, to *simply evaluating the implicit/inherent cognitive features presenting themselves in an action*, compliment what Randolph Clarke further denotes, while giving his own description of Quality-of-Will Theory: “. . . in blaming agents we [hypothetical Quality-of-Will theorists] are responding to what we take to be ill will or indifference manifested in their [agents'] conduct.” (246).

Now to look at Finn once again, with the intuition specified and the underlying account rudimentarily fleshed. Clearly there is no *ill-will* in refusing to return an escaped slave. Rather, it is the inverse impulse and conduct that seems to imply and/or “manifest” *ill-will and indifference* towards another (i.e. for one to return Jim to his owner would clearly show a great indifference to another's well-being). One obviously ought not to be subjected (nor ought one subject another) to such abhorrent cruelty as slavery in the American South encompassed. So it seems reasonable to infer, that some morally good cognitive feature of Finn's might be manifested in his morally appropriate conduct, and/or might be implied by his compelling aversion to what he thought to be ultimately right (i.e. to return Jim). His conduct seems (nonetheless) *very much praiseworthy* and it would be intuitive to say that it is morally commendable to have the aversion which Finn succumbed to. That being the case, Quality-of-

Will seems, at least in the case of Finn, to be superior to the type of akratic account we've explored above, which would un-intuitively demand that we blame Finn for doing what he believes to be wrong (i.e. helping Jim escape). This, of course, is a rather absurd result, especially considering the morally relevant circumstances.

However, it's important (as with the Akratic Account) to as concretely as possible show the primary emphasis in the Quality-of-Will account of moral responsibility. So, in presenting greater specification for the Quality-of-Will account, there is (in contrast to the Akratics) held to be a discernable manifestation of *some morally salient feature of the agent's own self* in an action, which Quality-of-Will Theory sees as sufficient grounds for moral evaluation. It doesn't matter (for the sake of moral evaluation) if one is/was *aware* of the ultimate (or even subjective) wrongness of their action(s), nor does it matter if the action is/was (or not) an expression of one's intentional *control*. Rather, the agent's conduct is seen to *implicitly contain content that is morally salient, in being the type of action that expresses something about the agent (themselves)*. However, when actions have this crucial connection and which feature(s) of an agent's are morally evaluative (and which type and strength of reactive attitudes we might be entitled to respond with) differs per account. But so as to (at least) fully sketch the account at hand (i.e. the Whole Self Account) and so as to see how an account within the domain of Quality-of-Will structures itself explicitly, it will be helpful to see which specific features are relevantly morally evaluative (and how so) in the Whole Self Account specifically.

**a.) The Whole Self Account Specified:**

For Arpaly and Schroeder it is said to be psychologically “*well-integrated*” beliefs and desires which properly arbitrate praise/blame attributions (175). So, accordingly, if Finn’s well-integrated beliefs and desires had (conversely) been in support of bringing Jim back to his owner, we would (according to Arpaly and Schroeder) advance blame against him. But also, should his beliefs/desires not have been well integrated, even though the same outcome resulted from Finn’s actions (hypothetically), the degree of praise is oriented by the degree of integration (Arpaly and Schroeder 174). So we would blame/praise Finn in accordance with a given morally salient feature’s incorporation with the totality of his psychological states and dispositions (e.g. how it integrates with his aversions, rationalizations, personality, etc.). Importantly, there must also be “some very minimal standard of integration” for evaluation of the action to be attributed to Finn at all (Arpaly and Schroeder 175).

Now, specifics of the *Whole-Self* account aside, what can be said of Quality-of-Will (henceforth *Q-of-W*, for brevity) accounts on the whole (i.e. in general)? Well, as should be clear at this point, if positively morally evaluating *the inverse of akrasia* is plausible, then akratic action cannot be the *sole* proper grounding for moral responsibility (and/or for arbitrating praise and blame attributions). Instead, the account serves to reject entirely the necessity of the akratic, Zimmermanian conditions for moral responsibility laid out above. It shares this important feature and emphasis with various other Q-of-W accounts (as will be seen). And discerning the shared basic structural features between such accounts will be instructive in further understanding the *broader theoretical framework underlying all of Q-of-W*. Indeed, already this supervening structure has shown itself. For if the *Whole-self* account can properly

be labeled as Q-of-W (with such a profound rejection of *both* Zimmermanian conditions) this same opposition to akratic conditions for moral responsibility should (at least) be reflected in other accounts. Taking a couple brief examples below, it seems the trend holds across accounts. Although, of note, the accounts used below are technically a subset of Q-of-W, known as *Attributionism*, though the specifics of this sub-theory are irrelevant for the current purposes of showing the super-structure to which all adhere.

### **b.) A Universally Shared Rejection among Quality of Will Theorists:**

As far as the (akratically based) epistemic condition that agents *must be aware of the wrong-making feature(s) of their actions to be blameworthy (culpable) for them*, Q-of-W theorist Mathew Talbert expresses explicitly the contrary in saying: “. . . an agent who never acted wrongly by her own lights may still be open to moral blame.” (48). Of course, the very same contention is shown in the Finn case above: Finn thought he acted wrongly, but received praise despite this fact. Similarly, the control condition in akratic evaluations (whereby one must have been allowed to exercise volition and choose an alternative) is eschewed entirely as well by Q-of-W. Take Angela Smith’s *Rational Relations View (or for brevity RRV)*, which directly rejects volition and choice in sourcing moral responsibility: “. . . what makes an attitude ‘ours’ in the sense relevant to questions of responsibility and moral assessment is not that we have voluntarily chosen it or that we have voluntary control over it, but that it reflects our own evaluative judgments or appraisals.” (237).

Now, with Smith’s RRV and the Whole Self account, we see a shared rejection of the control condition, but we also see an important dissimilarity between them in that, Finn was compelled

by an affective aversion to act contrary to what he *rationalized* to be the best course of action, yet his being compelled to act contrary to how *he rationally ought* (i.e. towards the end of his reasoned moral goal, to return Jim) is what is morally evaluative, what is praiseworthy in the Whole Self account. With RRV, however, this same moral evaluation would seemingly not be possible because it tracks agent's ultimate reflectively arrived at "evaluative judgments" and "appraisals" instead. However, it's even more useful to see where dissimilarity arises from Arpaly and Schroder's Q-of-W account and so as to restrict the scope of any universal claims we may make about the nature of Q-of-W. Accordingly, it should be noted that *irrational actions may not be universally morally evaluative objects for Q-of-W*, but *control* is seemingly not necessary for moral responsibility on either account.

### **c.) Two Base Commitments of Q-of-W:**

Admittedly, I'm risking unifying multifarious accounts, which differ in some *not* insignificant respects, but hopefully the explicated conditions that follow have been adequately demonstrated. The two commitments ultimately stem from rejection of the conditions for moral responsibility established earlier. Although, it must be noted, this approach is to take a page from the already established precedent set by Jan Willem Wieland, who similarly attempts to make universal claims about Q-of-W; so for the purposes of this work, it will be helpful to note her yet further remark that, ". . . they [Q-of-W Theorists] deny that blameworthiness for unwitting acts is to be explained by blameworthiness for benighting acts. Instead, blameworthiness for unwitting acts is to be explained by lack of moral concern." (Wieland, 19).

So following Wieland's lead (although without presenting the full extent of her enumerative theoretical scaffolding), *what might be some universal commitments of Q-of-W, in relation to the analysis thus far?* Well, let's stipulate two *prima facie appropriate universal commitments for Q-of-W*, which seem plausible from our examination of Arpaly and Schroeder's case, in addition to Wieland's contribution to this end, and abstracting from the elements of the two tertiary *Attributionist* accounts briefly cited above:

- 1.) *Q-of-W rejects the necessity of either awareness and/or control for establishing agents' culpability and/or in attributing praise/blame.*
- 2.) *Q-of-W accepts that agent's genuine moral character, and/or genuine moral commitments, can be manifested by (and rightly inferred from) agent's actions and/or omissions (at least of the relevant kinds).*

## **5. Agent/Act Disconnect and Difficulties in the Whole Self Account:**

With this in mind let's return once again to the *Whole Self* account; however, let's move past Finn for a moment and view another case in Arpaly and Schroeder's piece, regarding potential instances of "*out-of-character acts*" (182). The case, I think, will be instructive in seeing the nature of this sometimes elusive and contentious connection, between an agent and their action. The difficulties to examine arise quite prominently in the example of *the uncharacteristic murderer* (as I'm calling their case). The case is mercifully brief and simple in its presentation.

“A person who seemed decent if boring is convicted of first-degree murder” (182).

However, the complexity comes in analyzing how we should evaluate the murderer, given these circumstantial stipulations: **a)** Nothing about the person leads one to think they are of such a will or character so as to be capable of this blatantly morally infringing act, in fact, contrary to this impulse, the person seems otherwise “*decent*” and morally upstanding. Yet **b)** the person (inarguably) committed the offense, engaged in an action which we want to think morally depraved persons engage in. These two facts do not coincide well with one another in the Whole Self account; for being able to commit murder does not seem within the purview of a morally decent person. So how do Whole Self theorists reconcile past judgments of a person’s former decency with their recent crime of murder? Arpaly and Schroeder posit that, “Regardless of what one thought of her in the past, the knowledge that the person has committed murder is a reason to conclude that she has a personality or character with which the act of murder is reasonably well-integrated. Thus, she receives full-blame” (183). However, they also note that, “Such full praise and blame [praise regarding an inverse case not presented] are not given when there are special reasons to believe that the person *is* what she always appeared to be and the act *was* poorly integrated, after all” (Arpaly and Schroeder 183).

So, according to the Whole Self account at least, an act can give us *prima facie* reason to believe a person is (or is not) of a certain genuine moral character. Knowledge of one committing murder gives us reason to think they have a character that corresponds (or connects) to *genuine ill-will and disregard for others*. It then seems we have some reason to reject our previous thought of the person’s moral decency and their tame or boring nature.



Nonetheless, this case examination serves to show that moral evaluation is arbitrated solely through an action's proper "integration" within an agent's whole-self on this particular account. This is the type of connection we specified as being needed for any moral responsibility account and now we see at least one enumerative example on full display. But how does this particular account (as one among the many Q-of-W accounts) show a disparity with Akratic accounts, regarding what connections must exist between agents and actions?

Well, the Whole Self account does make one contention that seemingly all Q-of-W theorists make about action/agent connectedness, namely: *that actions are interconnected with agents in such a way as to provide morally evaluative content concerning the acting agent* (though the Whole Self view is tamer than others in their respective camp, in saying it can be later undermined). In the presented case above, the fact that the murder occurred was enough to warrant *tentative* blame. We didn't first have to ask something like: was the agent aware of what he was doing? Or aware that what he was doing was morally wrong? Was he coerced into committing the murder? Or criminally mentally-ill such that he could not understand what he might be morally expected to do? These questions are necessary first steps for akratic theorists, before thinking any act tells us something morally evaluative about the actor. So this case has shown us (at the very least) that there is importantly not parity with *prima facie reason-giving import for unqualified actions in akratic accounts of moral responsibility*, such as the Whole Self account permits.

But beyond merely the sufficiency or insufficiency of *prima facie* reasons in moral evaluation, if we want to establish substantive connections between actions and agents, so as to justify

them as even being proper targets of our moral evaluations to begin with, we would need to be certain (should it somehow be definitively established) that someone's possession of ill-will manifesting cognitive features *actually entailed one's blame* (or even more loosely, their *moral evaluative accountability*). Regardless of the difficulty of trying to show when some character trait or disposition is genuinely manifested by an action, it seems obvious that we're begging the question, unless we first establish that features such as these have any morally evaluative content to begin with. And, indeed, Rosen provides us with a rather striking counter-case, which leads us to think that merely attributing morally bad traits or cognitive features to an agent is not a sufficient condition for holding agents culpable and blaming them. The connection between agents and actions, which would be sufficient culpability and for attributing blame and praise must additionally firstly know *how these cognitive features or characteristics came about*. This will be shown to be an equally important part of the story, in figuring out when (and how) we can morally evaluate someone for some action/omission.

## **6. Blamelessly Bad Bonnie: A Counter-Case for Q-of-W**

Let's now look at a case of Rosen's that will demonstrate the true necessity of knowing the origin of one's morally evaluative cognitive features or traits, since the agent in his case both: **1)** clearly displays and genuinely possesses morally evaluative cognitive features or traits, which she straightforwardly manifests in her actions. Also, **2)** she appears entirely *blameless* in coming to possess (and eventually manifesting) these features or traits. A condensed and rough reconstruction of the case of Bonnie (from "*Culpability and Ignorance*") follows:

### **a.) An encounter with Bonnie:**

You are standing on the curb, during a heavy rain, with your children. You're all utterly soaked to the bone, when you finally (after many failed attempts) manage to hail a cab. But as you enter the cab in comes Bonnie, who runs over from across the street and violently pushes you and your children out of the way, so as to steal your cab. Immediately, you blame Bonnie for what seems an egregiously (even pathologically) selfish and inconsiderate act. But you later have the serendipitous chance to confront this woman at a party you're attending. You initially attempt to charitably give her every possible exculpatory out (i.e. asking if the action might be justified by, ". . . duress, necessity, ignorance, or diminished capacity" [Rosen 77]), however Bonnie flatly refuses any and all possible excuse. In fact, the only explanation Bonnie gives is: "When I saw the cab pull over, I weighed the costs of inconveniencing you against the costs to me of waiting for another. In the end I decided that it made most sense for me to take the cab" (Rosen 77). Rosen summarizes the initial evaluative intuition perfectly, "Apparently this stranger is an unreconstructed selfish creep with no excuse whatsoever for her conduct." (77).

There are important further details to Rosen's case, some of which will follow, but we now have the initial setup and it is worth stopping to examine just how Q-of-W accounts would respond (at this point), since what is specified is enough for Q-of-W to justifiably perform some moral evaluation of Bonnie (i.e. her actions imply morally objectionable cognitive features and she even expresses a profound ill-will and indifference to you and your children's plight). Intuitively, it seems like we needn't specify any of the Akratic conditions for moral responsibility in establishing her blame. Bonnie rather straightforwardly warrants a negative moral evaluation (and one's reactive blame is hard to deny). Perhaps showing agents meet the Akratic conditions

constitutes actually unnecessary steps. Well, let's take in turn some Q-of-W accounts and explore the sufficiency of each's moral evaluation of Bonnie.

**b.) Whole Self:**

For Arpaly and Schroeder, it seems that the morally salient features of Bonnie's character are reasonably well-integrated with her action. She doesn't feel some contrary affective aversion to pushing you and taking your cab, nor is she the slightest bit conflicted in the deliberation that motivated her conduct. She even endorses fully and reflectively the fact that she simply wasn't willing to wait for the next cab, despite that decision entailing harm to you and your children. Her actions are entirely in character. Accordingly, Bonnie's blame is, to the fullest extent, commiserate with the wrongness of her action. The contents of her character, and psychology, are wholly amenable to the objectionable qualities of her action.

**c.) Rational Relations View:**

For Angela Smith, we see that Bonnie's attitudes towards others correlates with her rational and evaluative judgments perfectly. That is to say, Bonnie admits (without qualification) that she engaged in clear-eyed, rational deliberation and then acted in accordance with the outcome of this deliberation. And so as to ensure we are tracking Angela Smith's account of moral responsibility properly, it should further be specified that Bonnie is premised as being in possession of *fully intact rational capacities*. So it is not some type of unfortunate defect in her ability to reflect and/or reason properly, which led her to misguided and immoral commitment, quite to the contrary, Bonnie has adept faculties of reasoning (see Rosen 78). Her "*evaluative*

*assessment*” and “appraisals” (see A. Smith quote above), then must show a glaringly objectionable disregard for others, and her action is rightly seen as a pure manifestation of her moral commitments.

Again, it’s explicitly premised by Rosen that Bonnie weighed the normative force of her own desires to get out of the rain, against the normative force of considerations for a soaked stranger and her children, yet she was simply unconvinced that she needed to wait for the next cab. Bonnie’s evaluations, attitudes, and judgments are laid bare before us and they are plainly selfish and in the absence an exculpatory defense, we blame her unrestrainedly for her actions.

Seemingly, so far so good for these two accounts. Q-of-W at least appears to be following our pre-theoretical intuitions properly. If the Akratics need more circumstantial specification to reach moral responsibility and/or to advance blame, then the Zimmermannian conditions seem extraneous. Bonnie is premised as acting as an indisputably “*selfish unrepentant creep*” (Rosen 78); so a theoretical account that warrants blame accordingly, without preconditions, seems eminently more preferable. Or so at least it would seem. However, the next part of the case is where the dispute between the Akratics and Q-of-W comes to a head and (in fact) the tables turn once again. For Q-of-W has already found it sufficient to blame Bonnie (given circumstances as they are), yet it will be shown that Bonnie is in fact blameless (as the case progresses) and so Q-of-W misses the relevance of this type of information in moral evaluations (fundamentally). Consider this further development in Rosen’s case:

**d.) An encounter with Bonnie’s Neurologist:**

Upon an occasion of further serendipity, it turns out that Bonnie's neurologist is present the party you're attending and overhears when you interrogate Bonnie. He approaches afterwards and then informs you that Bonnie has been stricken with a virus recently, that she used to be morally upstanding and otherwise decent, but that, ". . . the virus 'rewired' the neural circuits underlying her normative sensibility: her view of what matters, and in particular her view of what counts as a reason for action." (Rosen 78). This consequence is the sole effect of the virus upon Bonnie though. Importantly, she still, ". . . possesses the same general capacity for reflection and self-control as you or I." (Rosen 78). She also subjects her new "rewired" views to epistemically appropriate critical investigation, fully knowing the illness has influenced her new moral beliefs, yet they remain unchangeable. Her neurologist has even gone through exhaustive moral investigation with her as well ("from Plato to Korsgaard" [Rosen 78]) in an attempt to reverse the effects of the illness; yet she is not any further convinced of her old views on morality and the normative weight of her old beliefs simply does not obtain anymore.

Now we have an apparent problem for the Q-of-W accounts we evaluated Bonnie by earlier. For both accounts made their moral evaluations of Bonnie with the circumstantial information (as was) and while they could (potentially) easily still modify their assessments, it already goes to show that both accounts didn't see *the necessity in doing so*. This fact shows that the accounts are too narrow in their scope of sufficient conditions for attributing blame. Remember, the initial setup was shown to be enough for both accounts (RRV and Whole Self) to establish Bonnie's blame her for her conduct.

However, as Rosen says of a condition (in this case ignorance of one's wrong doing) resulting from disease: "[it's] the paradigm case of a blameless condition." (79). How can one rightly blame Bonnie for her virus altered "normative sensibility"? It would be (loosely) akin to blaming an Alzheimer's patient for not being able to remember his family members anymore. Although, normally, lack of consideration for one's intimates is morally salient and could potentially lead to justifiable blame, it certainly cannot play an morally evaluative role for someone afflicted with a disease that changes the nature of *the very thing being evaluated*. With Bonnie, it's strikingly parallel, since her new lack of properly attributed normative weight to the considerations of others is entirely brought about by the effects of her virus.

## **7. The Blameless Condition Principle Examined:**

However, let's pause to investigate the claim that Bonnie's condition is truly blameless. Perhaps some might want to push back at the notion. Though (as in the Alzheimer's case) it seems somewhat intuitive, we should look at and motivate a further contextualized defense of such a central notion in the case of Bonnie, so as to ensure the principle (of blamelessness emanating from disease) truly holds.

### **a.) Persistent Epistemic Duties:**

To begin our critical investigation, Rosen helps further contextualize the principle, soon after giving it: "To be sure, the fact that she has been transformed by the illness imposes special obligations of reflection." (79). One's diseased condition might not always exculpate, by Rosen's own admission, since we might (in virtue of our disease) have new epistemic obligations of

critical reflection. To help further motivate the concern, we can make a roughshod case to exam its necessity. Certainly, if one were *aware* that one's recently contracted head fungus caused one to have extremely vivid hallucinations, one shouldn't just continue to uncritically accept one's sensory perceptions from now on. Indeed, if one were to continue to do so and to take their presenting reality as veridical and as not in need of special critical reflection, we might well blame them for any resultantly morally bad conduct.

To give a more specific case: say this fungus-afflicted person hallucinated that a monster was knocking at their door, right around the same time as they were expecting the pizza they ordered to be delivered. Despite the many hallucinations they had experienced earlier and despite their knowledge of how their head fungus affected them, they still went ahead and attacked the apparent beast with a kitchen knife (not taking a second to question the strange occurrence and perceptual appearance). Once the hallucination subsided, they came to realize it was only the pizza delivery man they had slain.

We naturally want to hold this person accountable and ought to negatively morally evaluate specifically *their failure to be appropriately critical*, given their foreknowledge of how their recent disease (i.e. head fungus) affects their perceptions. Rosen's concern seems important to consider and it ultimately shows that one's blamelessness is still subject to certain epistemic obligations (i.e. "special obligations of reflection" [79]), even when one's cognitions are affected by a given disease. Certainly, he's sure to specify of Bonnie that: ". . . [She] does not acquiesce uncritically in the normative appearances; she is thoroughly critical, and her views



survive this criticism.” (Rosen 79). So Bonnie, unlike our hallucinating subject above, clearly meets her epistemic obligations.

**b.) Negligent Contraction:**

One more potential constraint is that one *not have been negligent in contracting one’s disease*. Although it’s a bit of far-fetched case revision, I posit there are circumstances in which one could bring about one’s disease (and resultant diseased condition) in blamable/culpable ways. To give a case towards proving this contention:

Suppose it turns out that Bonnie’s particular virus is commonly well known and scientifically well understood; further yet, its only means of communication are easily avoidable. In fact, the virus is only contractible through eating expired pufferfish from the Sea of Japan. Bonnie, on her recent trip to Japan, stumbles into a run-down sushi joint looking to eat her favorite dish: pufferfish. The chef of the establishment cautions that the pufferfish looked off and that he thinks it was from a previous shipment. However Bonnie insists he prepare a fillet of it for her. The chef hesitates to prepare it, but Bonnie persists and waves huge amounts of cash in front of his face, she even threatens to negatively review the establishment online, contending she is a prominent food reviewer. The Chef acquiesces to her threats, fearing for his fledgling sushi bar. Bonnie can unmistakably smell the off notes of acidity in the fish when served and the chef warns her it can be a vector for a specific virus when rotten. Bonnie knew this already, but always wanted to eat authentic pufferfish sushi and she wasn’t going to be in an authentic Japanese sushi restaurant again anytime soon. She knows what the symptoms of the virus are but eats the putrid fish anyway. She’s stricken with the virus soon after.

So, it seems if one is negligent in contracting a disease they similarly don't maintain blamelessness for effects of it (at least if one knowingly risked contracting it, and did so freely). So there might be this further step in clarifying when a disease-brought-about-condition is truly exculpatory. However, it's not hard to imagine that Bonnie didn't do some particular action to bring about her disease (certainly nothing in Rosen's case leads us to think she did), and many diseases (of course) don't even have such avoid-ability. But for those that do, we might need this further condition. However, to be clear, none of our investigations have precluded the blamelessness principle in any substantive form, so it seems fair to posit that as a whole, disease can be an exculpating condition.

But what types of responses are we justifiably entitled to in the case of Bonnie? What sense of responsibility is there in cases of disease altered "selves"? How about the targets of Bonnie's (still unchangingly reprehensible) conduct: are they not entitled to negative evaluations of her? Are there problems of identity in cases like Bonnies, which are interfering with blame/culpability attribution? The questions that arise from Bonnie's case and our analysis thus far are not easy to answer. We have vacillated between both Akritic and Q-of-W accounts of moral responsibility now and basic questions remain unanswered. Perhaps Zimmerman will once again have some useful distinctions in furthering our examination. Let's start our honed investigation on notions of "self" related to moral evaluation. The case of *Anne and Bob* and the notion of the "*Not-Too-Shallow Self*" will be instructive in seeing some of the issues emerging with the case of Bonnie (see, Zimmerman 229).

## **8. Locating Morally Evaluative Selves:**

The case of Bonnie has shown us that it is not enough (for the sake of moral evaluation) to merely demonstrate an agent's possession of some particular features, or qualities, that are (in themselves) morally objectionable. The morally evaluative interconnection between these features or qualities of an agent and their resultant actions can be severed if the features or qualities are brought about in some blameless way (as in the case of being merely symptomatic of some unavoidable disease). But when exactly then are we morally responsible (and blame-/praiseworthy) for our actions and for our possession of reprehensible features of cognition or character (e.g. our attitudes, beliefs, desires, etc.)? We need to be able to identify a sort of primary, genuine identity and then show that something morally salient is attached to it in a fundamentally culpable way. However, the case of *Anne and Bob* (and some further revisions thereof) show that delineating and locating evaluative "selves", in relation to which one's behaviors and features might be properly interconnected, is a difficult prospect. The brief and rough reconstruction of Zimmermann's case is as follows:

**a.) Bob and Anne:**

Both Bob and Anne utter the same morally reprehensible racial slur. However, unlike Anne, Bob suffers from a form of Tourette's syndrome, which manifests verbal ticks.

So, initially, we can plainly contend that the action itself (uttering a racial slur) is equally morally reprehensible in either iteration (i.e. when uttered by Bob or Anne). But again, as with Bonnie, specified circumstantial facts and agential conditions circumscribe our initial evaluations of the agents themselves, in relation to their behaviors. Once we learn that Bob suffers from a condition which could (let's say it in fact did) manifest the morally objectionable

behavior in question, he then is seemingly excused/blameless for it. As Rosen contends in the case of Bonnie: disease is *“the paradigm of a blameless condition”* (see prior quote above). So then Bob’s exculpation seems easily enough understood. But why might we continue to blame Anne? It’s of course specified that she is not afflicted by a disease, which could be manifesting her behavior, but nonetheless given further information, might we also come to excuse (or at least not blame fully) Anne?

As Zimmerman motivates, what if Anne’s uttering a slur was merely a slip of the tongue? What if, “. . . though harboring racist tendencies, she is committed to rejecting racism and typically succeeds in overcoming her inclinations, but this time the slur just popped out—” (229)? Do we now need to also say that Anne is excused, given these further specifications of her character, commitments, and attitudes? Well, that is precisely the question, which seems very much open to debate.

In one sense, the Akratics would be quick to point out that Anne did have greater volitional control than Tourette’s afflicted Bob did (although, to an extent at least, I suppose some might press that such slips of the tongue are actually involuntary; however, for present purposes let’s suppose there is some volitional control in slips of the tongue). So in having control over her utterances and in being aware that racial slurs are morally bad things to say, Anne appears morally responsible to Akratics. Although, in another sense, Q-of-W theorists may want to say something along the lines of, “the fact that such a reprehensible slur could even come to mind in the first place and further yet actually slip out of one’s mouth, serves as proof enough of

one's morally importune prioritization of concern for others". Perhaps we ought to blame her for the objectionable features of her inappropriately prioritized concerns.

However, as with Bonnie, to try and construe our moral evaluations along either of these lines is to entirely miss the point. How did Anne's cognitive features (like her "racist tendencies", which we find morally objectionable in and of themselves) come about? The refined question is: *how can we say that Anne (or even Bob) is truly interconnected with their presenting behavior, in such a way that we can say the negative moral evaluations we want to advance (or conversely withhold) are justified/merited?* To more clearly see the issue at play, we can revise the case and show the interconnection between agent and behavior to be even more tenuous than in the initial case.

**b.) Bob the Racist (revision #1):**

As it turns out, Bob is a racist in the fullest and most robust sense. He attends Neo-Nazi rallies every weekend and very often intentionally utters slurs directed at various minority groups. He also does (as before) have Tourette's syndrome. He utters a racial slur today, but it just so happens to be entirely an involuntarily tick (this time at least).

What now are we entitled to say? Perhaps the utterance is still blameless. After all, a blameless condition is at play (i.e. a disease). Although, maybe the appeal of the Whole-Self account shows itself well in this context. For we could say that (even though potentially incidentally, involuntarily uttered) the uttering of racial slurs, in general, is very much in accordance with Bob on the whole (i.e. given the action's befitting integration with his other

beliefs, desires, behaviors, and character attributes). We would have to say, however, that the Whole Self account is diachronic, for in the moment of instantiation the action has no evaluative content at all in being solely produced from a neurological disorder. Furthermore, there seems to be serious problematic implications *for any* account of morally evaluative selves if it were to incorporate temporally distinct aspects of self. For instance, why should the attitudes one held at fifteen have any bearing on the attitudes one manifests 20 years later? I don't want to dwell on the metaphysical aspects of "self" here too much, but it should be clear (and this is the only relevance for the argument at hand) that things get messy *very* quickly, especially when we need to specify such controversial properties of identity, in order to maintain an account of moral responsibility. Admittedly, this insistence upon precaution does not (in any dialectic sense) defeat anything; but regardless, it still is relatively clear that some behavior's being brought about by a syndrome (at least when not negligently contracted at some previous point and when one has met any special obligations in virtue of being aware of their diseased state) does not confer anything of morally evaluative import upon the agent in question. The utterance itself is still obscure as an object of moral import. From whence does it originate and how is it interconnected with the agent themselves? This question only becomes more and more pertinent.

**c.) Bob the Indecisive (revision #2):**

Taking a yet again revised Bob, let's say he is the perfection of indecisiveness and is (quite literally) entirely split when it comes to deciding on the proper normative weight of saying racial slurs. Imagine that Bob's desire for tolerance and hatred exist in totally equal proportion.

If one could possibly be thusly composed, uttering a racial slur would equal parts conform to his genuine desires and also contradict them. Importantly, let it be stipulated that these desires are as high-order as possible, and to every extent reflect his genuine desires (given whatever criteria one needs to understand the genuine desires by).

Are we to imagine that Bob has a disparately bifurcated self, which exists in two separately evaluative ways (i.e. each analyzable of its own accord)? Can we blame Bob A (the hatred dominated cognitive component), while praising Bob B (the tolerance dominated cognitive component)? Clearly, this is an entirely absurd result. Bob cannot possess within him two independently morally evaluative selves. Either Bob (simpliciter) is, or is not, connected to his slur uttering in a minimally culpable way. We cannot contradictorily contend that Bob both is and isn't blameworthy. But again, why does it matter what we say here? We would have to first know **1)** how Bob's desires came to be the way they are, and then **2)** which desire is being occurrently manifested in the behavior itself. If Tourette's syndrome truly compelled him to utter the slur, we wouldn't want to blame at all.

The concerns that arise from the revised cases of Bob are perhaps multifaceted, but primarily it brings into question whether it is fundamentally possible to have a precise enough account of moral responsibility, such that it could properly discriminate between a "not-too-shallow-self" and a ". . . self that would seem to disappear altogether" (Zimmermann 229). Where one's self is too "shallow" actions attributed to it (or perhaps emanating from it) would lack morally evaluative import; i.e. it represents the limits of the "*worthy*" self, the self that could warrant attributions of praise or blame. Where this self disappears, moral responsibility

disappears when traced to it. So when nothing is imputed to an agent by an evaluation of his/her actions or features, here the self has properly vanished. Every account of morally responsibility must contain this important delineation and a lot hinges on how it is fleshed out.

As Zimmermann's elucidates (and problematizes), so called "*Deep Self*" views run this fundamental risk of over-extending and/or undercutting the limits and depths of properly morally evaluative selves (see 229). Taking a look back at the Q-of-W accounts we've explored thus far, each seemingly has an adherence to some form of a "*Deep Self*" view. In the Whole Self account, the deep self is the one that is well integrated across all cognitive strata (for instance, the self that emanates holistically from one's rationalized beliefs, affective aversions, character traits, etc.). The totality of integration is the connecting strand between the various cognitive features contained within oneself and only when weighed in totality can any particular aspect be imputed to *the actually deep self*. As we saw previously, this is of course how Arpaly and Schroeder determined an appropriate object of moral evaluation. Similarly with Smith, we saw that the deep self is maintained in (and limited to) the rational reflections and the evaluative commitments that arise from one's reasoning about a given moral matter. Only entailments of one's evaluative judgements or commitments, can constitute the morally salient aspects of one's truly deep self.

Importantly though, how do these accounts specify the "*too-shallow-self*"? The cases of Bonnie and Anne and Bob have shown that an action's (or cognitive feature's) connection to a sufficiently deep (i.e. an appropriately morally evaluative) self can come apart in dramatic and non-anticipatable ways. How would they respond to these cases? I'm contending: *not well*.



## 9. The Fundamental Problem for Q-of-W: Act/Agent Disconnection:

We've seen already that both RRV and the Whole Self account flounder with Bonnie; i.e. they prematurely form moral evaluations positing her blameworthiness/culpability, without the ability to account for the blameless ways in which she came to be the disease afflicted, and resultantly inconsiderate creep, she is. The reason that they failed to be able to exculpate Bonnie though has to do with something basic and inescapable. Remember the second established universal commitment of Q-of-W:

- *Q-of-W accepts that agent's genuine moral character, and/or genuine moral commitments, can be manifested by (and rightly inferred from) agent's actions and/or omissions (at least of the relevant kinds).*

With this commitment, there are going to be irresolvable problems in evaluating agents in cases like those that Zimmerman and Rosen give. Agents can act in ways *fundamentally disconnected from their moral deliberations and from their (non-disease afflicted) selves*. But this commitment of Q-of-W bases itself around a very tenuous interconnection, which certainly does not obtain universally between agents and their actions. Again, Q-of-W could modify its evaluation once the disconnection is shown, but it would have to do so only retroactively. If the disconnection between agent and action can be definitively discerned, then fine Q-of-W escapes complication; but as has hopefully been shown by this point, where actions and agents

connect or disconnect in properly morally evaluative ways can be extraordinarily difficult to discover.

Furthermore, how can the Whole Self view (for instance) respond to *Bob the Indecisive*?

There is an equally proportioned disposition towards (and commitment to) hatred and tolerance. Both morally salient attitudes are precisely equally integrated. It doesn't seem like the view can dismiss the utterance of a racial slur altogether, there is something clearly morally evaluative in the act, but to take the act to be necessarily connected to one's genuine (evaluative) self is certainly not going to cut it with Bob (unless they fully embrace the absurd metaphysics of a Bob A/Bob B dichotomy).

To further emphasize the problem for RRV, what about when one's moral deliberations are blamelessly influenced by a condition brought about by disease? When the appeal of other's interests and one's own have been muted and/or amplified, so as to change the otherwise (non-afflicted) commitments an agent's process of reasoning would conclude upon, we can't simply stop at identifying the objects of their deliberative and reflective processes. The processes can be non-culpably skewed, or disrupted in fundamental ways, such that products of such processes are *potentially purely incidental* and attributable to an agent in a superficially external way. As an example, the product of disease is not reflective of the deliberative machinations of the agent themselves.

### **10. The Problem for Akratia:**

While the Akratic accounts do avoid the *over-extension* of morally evaluative interconnections between agent's and their actions, they also *under-extend* and miss proper evaluative incorporation of cases like those of Finn, which showed us that there are morally evaluative elements of agents (e.g. heart/aversion) which cannot be properly captured within the domain of volitional and clear-eyed acts. Despite Finn's intentions and volitions, we praise him. His sub-conscious aversions are what we praise, his inability (not his ability) to act in accordance with his volitions and clear-eyed moral deliberations is *why we say he had a good heart in the end*. Perhaps we've reached a bit of an impasse then, both Akratic and Q-of-W camps seem to have inherent insufficiencies and problematic cases. Is there another type of view, which might be able to straddle a viable middle ground, i.e. not over-extending, nor under-extending in denoting the class of evaluative connections between agent's and their actions?

Well there is another aspect to moral responsibility and the morally evaluative connection between agents and actions towards which we might now finally look: maybe *commonplace normative expectations* are the key to looking beyond the problematized camps of Akratics and Q-of-W. Randolph Clarke's shift in emphasis, as will be exemplified in the case of **Carol and the Hot plate**, might just provide something neither Akratic, nor Q-of-W accounts, can properly track: a commonsense demand in moral evaluation, the demand that a person be accountable to *reasonable expectation*.

### 11.) Clarke and Reasonable Expectation:

Clarke's case of *Carol and the Hot plate* will attempt to demonstrate that even when one's ignorance is plainly blameless, one still may be blameworthy for one's resultant conduct. The view does so by contending that one can sometimes *directly* and *commonsensically* be blameworthy for one's "*unwitting wrongful conduct*", irrespective of *how* one's prior ignorance came about (though in a different way than in Q-of-W). That is to say, the priority is not in tracing the moment of previous "*cognitive impairment*" and trying to understand how one might be blameworthy indirectly for the conduct that follows. Rather, what matters for moral evaluation (Clarke contends) is simply the agent's meet, or fail to meet, a commonplace normative demand that emerges from our, ". . . commonsense views about the psychological capacities and abilities to act that people ordinarily possess," (234). So to begin with the presentation of Clarke's view, a brief iteration of one of Clarke's three central cases:

**a.) Carol and the Hot Plate:**

Carol goes about her daily routine of cooking breakfast on a hotplate one morning. There is nothing extraordinary about the circumstances of this particular morning; e.g. no emergency to which she suddenly had to attend, no sudden meteor storm which distracted her from her daily tasks, etc. Yet today (given no foreseeable reason), she forgets to turn off the hot plate. In the ensuing fire, which consumed their apartment, all of Carol and her roommate's possessions were destroyed. However, it's important to note of Carol that, "She is generally considerate of their [her roommates'] interests, she isn't a careless person, and she isn't generally forgetful" (240). Nonetheless, irrespective of Carol's ignorance of having left the hot plate on, and given her roommates' awareness of her generally appropriate moral sensitivities, her roommates

hold her blameworthy for the fire and they are (perhaps rightfully) upset with her over their immense and ultimately preventable losses, which resulted from the hot plate fire.

But what exactly does this case tell us about Carol's blameworthiness? There is certainly an intuitive sense in which she might be morally responsible for the harm caused by her actions, but what specific feature in the case of Carol is behind her roommates' and our own readiness to hold her accountable for the fire? Well firstly, it's worth noting that if we are to morally blame Carol (directly, that is) for not turning off the hot plate, it would be to focus our moral evaluation on an omission, of a particular kind, which is also prevalent in Q-of-W accounts. However, the strength and source of morally evaluative import is seemingly entirely different in Clarke's case, than it is in an account like Smith's RRV (as an example).

### **b.) RRV and Clarke on the Moral Import of Forgetting:**

For instance, in Smith's seminal piece on RRV ("Responsibility for Attitudes: Activity and Passivity in Mental Life") she even begins her discussion with an instance of forgetting (i.e. one's forgetting an intimate's birthday), which is shown to correspond to an importantly different normative demand than is at play in the case of Carol. For Smith the demand surrounds *that one prioritize considerations for one's intimates* (or those of significance in one's life), so as to be able to remember the important dates in their lives. Smith wants to say of forgetfulness that it's relevant for RRV in the following way specifically: ". . . if one judges some thing or person to be important or significant in some way, this should (rationally) have an influence on one's tendency to notice factors which pertain to the existence, welfare, or flourishing of that thing or person" (244). She takes it to be straightforwardly indicative of one's

underlying moral character, of one's genuine moral commitments to the well-being of another. Moral evaluation would follow accordingly, making use of the appropriateness of one's prioritization of attention given to the interests of one's intimates (or conversely lack thereof). But no such implications seem to be imported in the moral evaluation at play in Clarke's case of Carol. Naturally, it's even specified of Carol *that she is generally considerate* and it would seem very strange (intuitively at least) to claim that one's forgetting to turn off a hot plate somehow directly (or even simply straightforwardly) shows something of significance about one's genuine moral commitments, or moral character.

Clarke does think there is a distinct normative demand attached to remembering certain things, but instead of taking an instance of forgetting (at least in cases like that of Carol's) to reveal something about one's moral character, Clarke says the following, "If it was reasonable to expect these agents to be aware that their conduct was wrong, and if they were free to do otherwise, what is their excuse? My suggestion is they have none" (244-245). The normative demand relevant for Carol then is grounded in the fact that, we often take normal persons, with the relevant capacities and in commonplace circumstances, to be beholden to certain *reasonable expectations*.

### **c.) Clarke's Conditions and the Explanatory Power of Reasonable**

#### **Expectation:**

Clarke's quoted bi-conditional above sets up concisely the two specific conditions in his account, upon which meeting, agents can be blamed; that is, he contends, when we find that: 1) *an agent was "free to do otherwise"* and that 2) *"it was reasonable to expect" the agent "to*

*be aware that their conduct was wrong”, then the agent can no longer be exculpated for their actions, even when purely unwitting. The first condition is directly parallel with the control/volition condition of moral responsibility we’ve seen before; however, the second condition is importantly different than the awareness condition in previously explored Akratic accounts (though it seems perhaps initially similar). It is different than the mere awareness of the wrongness of one’s conduct (at the time acting), which is the concern in the initial Zimmermannian condition of awareness we posited before. The second condition in Clarke’s account, more precisely, concerns instances when it would have been reasonable to expect an agent to be aware of the “moral significance” (241) of their conduct. That is, it’s not just that an agent was, or wasn’t aware, but *that they reasonably should have been.**

But, as seems fair to ask at this point, why is simply an appeal to the expectations we might commonsensically have for agents like Carol an appropriate basis for moral evaluation? Indeed, haven’t we seen that being too quick to adhere to our intuitive evaluations can actually lead to pre-mature attributions of blame/praise (as in the case of Bonnie)?

While that may be so in the case of Bonnie, Clarke gives us some convincing cases in which we seemingly can’t arrive at the correct evaluation without his appeal to what is reasonable to expect of others. To show the actual insufficiency of other accounts of moral responsibility to account for Clarke’s cases, taking a look back at Carol, how would we otherwise maintain her blameworthiness (i.e. given the previous accounts of moral responsibility we’ve explored)? There simply aren’t the means to do so, I would contend. For instance, the Akratic account would point to the facts that: 1) Carol was in control of her actions in leaving the hot plate on,

but that 2) she was not aware of the wrongness she was committing in leaving for work that day. She simply didn't realize that she had risked the apartment and her roommates' well-being, in leaving that morning. So we have not (yet at least) met the requisite conditions for blaming an agent (on an Akratic account). The Akratics would specify, that if Carol were culpable in her ignorance of leaving the hot plate on, then we might have grounds for the type of moral evaluation (like blaming), we initially see as appropriate.

But Clarke is clearly explicating in his case that Carol has not been negligent, nor somehow akratic, at some previous time; she *simply forgot* (as it seems intuitive to think persons occasionally do). We are blocked within the confines of an Akratic account from anything but exculpation; however, the intuition of ours and Carol's roommates' is to blame. So we need an appeal to something beyond the Akratic account. We can look to Q-of-W, yet it has already been shown that the instance of forgetting (at least in the case of Carol) is not of the same moral import, which of Q-of-W seems to demand (i.e. her forgetting doesn't illuminate anything about her deeper moral character, or moral commitments to other's well-being). It was simply reasonable to expect Anne to have turned off the hotplate, being that she had the capacity to do so (certainly, she has done it consistently many times before), and since there was nothing exceptional that directed her attention elsewhere from doing so. Essentially, if Carol possesses ". . . certain psychological capacities and abilities to act that we commonly think adult human agents ordinarily have." (Clarke 240), then we can seemingly blame her, since she didn't do what we expect agents like her to do (in the given circumstances). Clarke's account thus highlights one's "*unexercised capacities*" which we might *have reasonably expected one to*



*exercise* in a given instance in which they didn't (242). That is to say, our failures to manifest possessed capacities (be they cognitive, like paying attention, or something else) when we otherwise could have, can be a sufficient source of blame.

## **12. Responding to the Challenge: Clarke on Bonnie.**

But how does Clarke's "*commonsense reasonable expectation*" account (as it might be dubbed) of moral responsibility respond to the challenges levied against the other accounts we've explored herein? Taking up the case of an individual like Bonnie, perhaps there would be a couple different ways in which Clarke might seek to morally evaluate and classify such an agent. I think the intuitive starting point would be for Clarke to note that (simply) Bonnie is not subject to normal, commonplace circumstances. It doesn't seem controversial to say that Bonnie is subject to a very particular set of unique circumstances, which naturally do directly affect the way in which she morally deliberates. But it's not immediately clear how acknowledging this will factor into Clarke's account. There might be parity with some of what Clarke discusses in the case of *Anne*.

### **a.) Extreme Circumstance:**

Anne (to say it as briefly as possible) is one who fails to notice a stop sign and crashes into another driver as a result, but what's relevant about her case are Clarke's remarks about reasonable expectation and that there could be exculpating "*extreme circumstances*" (see Clarke 244). For instance, he notes initially of Anne that if a meteor were to have fallen a few blocks away, while Anne was coming up upon the stop sign, then running it might be excused

entirely, even though Anne had the capacity to be aware of it and was not restrained (in any particular way) from seeing it. So perhaps Clarke might try to circumscribe an agent like Bonnie by these means, noting an extenuating circumstance that creates an unreasonable demand, given said circumstance. But if Clarke's account merely relies on some set of exculpating circumstances as being relegated to the "extreme", it doesn't tell us much. When exactly are the circumstances appropriately extreme, such that we can exculpate agents violating commonsense reasonable expectations? It should be apparent that there are going to be disputes about which circumstances are *extreme enough to be exculpating*. For instance, meteor storms fit this category of exculpating extreme circumstance fine enough, but what of something like: a migraine, an acute moment of consuming existential dread, a breathtakingly beautiful garden on the same corner as the stop sign, etc.? All these types of examples distract from the task at hand (e.g. driving attentively), but they have varying degrees of normative force in exculpation (at least it seems intuitively). More would have to be said about the universal qualities of such exculpatory and extreme circumstances; there needs to be a delineation between the class of *distracting-yet-not-exculpating* circumstance and the extreme-enough-to-exculpate circumstance.

### **b.) Morally Non-evaluative Strong and Weak Psychological Capacities:**

Another way in which Clarke might be able to respond to Bonnie's unique set of circumstances is to note something about the change to the strength and weakness of appeals to her and other's interest. Clarke notes towards this end that, ". . . as a normative matter, what one can reasonably be expected to notice, think of, or remember on some occasion

depends in crucial part, on the extent and strength of the psychological capacities one possesses then.” (242). So perhaps Clarke could say of Bonnie that, given her rewiring, the expectation that she appropriately weigh the normative force of consideration for others becomes too stringent, since the strength of self-interest has expanded and the prioritization of the interests of others has diminished, as a side effect of the virus. She might just be excluded, given certain strength or weakness in her particular cognitive ability to weigh interests appropriately in moral deliberation (though importantly the threshold is not specified).

Rosen is insistent in mentioning that Bonnie still “. . . possesses the same general capacity for reflection and self-control as you or I.” (78). Thereby it is important to be clear that it doesn’t seem like capacity (qua capacity) is the type of thing that gets rewired in Bonnie, it seems like her *capacity* to morally deliberate remains intact. So might it still be reasonable to expect her to have come to the right conclusions? Perhaps, or perhaps not. But Clarke’s account can’t allow for consideration of the fact that disease *blamelessly brought about the change in strength of certain appeals of moral interest*. This is so for his account, since it fundamentally doesn’t appeal to how one’s ignorance comes about in attributing blame; remember, his account moves away from the morally evaluative import of benighting actions altogether and instead focuses on the capacities one has (at the given time and in the specific circumstances) to act, which we might normally expect ordinary agents to manifest. The normative demand comes from our expectations of these other ordinary agents with the same capacities, so how can we properly reference the fact that Bonnie’s moral deliberations are being influenced by a blameless rewiring? I’m not sure Clarke’s account can do so.

Is Bonnie potentially blameworthy for her poor moral deliberations, then? For her failure to meet a reasonable expectation to realize the wrongness of her conduct? Clarke's account worryingly seems to allow that she might be. For it can't be claimed that Bonnie is absent some capacity which others typically have; so in having that capacity appears that it is reasonable to expect that she be able to deliberate properly, since she could exercise her capacity, given that: 1) she had the capacity to be aware that what she was doing was wrong and 2) she was free to do otherwise (i.e. nothing external to herself was demanding she act one way or another, or be unable to recognize one fact or another).

**c.) Clarke on Finn:**

And what of Huckleberry Finn? Well, again there would have to be some account about *what would be reasonable to expect* for someone with Finn's specific cognitive capacities, the strengths or weaknesses thereof, and given his specific circumstances (e.g. having been raised in the American South during the widespread practice of slavery). I suppose, at this point, there is an apparent short coming in Clarke's account in that it seems to be assuming a commonsensical idea of when exercising a given capacity becomes unreasonably hard, due to its weakness, or in virtue of the strength of some other psychological motivating force. Also, the account assumes a commonsensical idea of when circumstances are appropriately extreme so as to override our reasonable expectations. Essentially, to truly understand if Clarke's account is robust enough so as to deal with the problem cases we've specified for other accounts of moral responsibility herein, it seems to demand he answer the prior questions concerning, when an expectation is in fact reasonable, given the various differentiations of strength in one's

capacities and when the normative demands of reasonable expectation obtain and when they are entirely diminished, to the point of exculpation.

While Clarke's account seemed promising, it appears to ultimately falter and the case of Bonnie, again, proves a foil for the account's central normative condition (i.e. reasonable expectation to exercise a possessed capacity). Although we have tried earnestly to resolve the problems brought about in the moral evaluations of agent's moral responsibility and blame/praise for acting ignorantly, it seems we are still plagued by a familiar set of problems and unanswered questions. In the final section, we will enumerate these and make a final tally of where hope for a sufficiently delineated account of moral responsibility might lie.

### **13. Conclusion:**

From the case of *Doctor S and Mrs. P*, to the case of *Carol and her hot plate* we have explored many different scenarios, in which we tried to spell out the appropriate parameters of moral responsibility for unwittingly wrong (or right) actions and to try to understand how agents fundamentally connect/disconnect from their actions, omissions, and cognitive features in both praiseworthy and blameworthy ways. However, it would appear as though not a lot of positive ground has been gained in our attempt, quite the opposite. We've come to see merely the insufficiency of multiple answers to the question of: *When are we morally responsible for our ignorance? And further: when and how are we blame-/praiseworthy for our actions and character/cognitive features?* Multiple established accounts from *Quality of Will* theorists, to *akrasia centric* views, and even (lastly) an account based in *commonsense reasonable expectation*, all seemed to lead to some unintuitive results, some absurd consequences, and/or

simply failed to delineate the boundaries of moral evaluation properly (or fully adequately). It's naturally fair to mention that our attempt to clarify such broad questions has been (by no means) exhaustive, there certainly are plenty of other candidate accounts, which might entirely respond to the problematic cases raised and which might answer the questions that remain unanswered here. Nonetheless, we attempted a gloss of two major camps in the literature (Akratic and Quality of Will theories), which seem to both be given to certain erroneous attributions of culpability and blame (*most notably in the cases of Huckleberry Finn and Bonnie*). We even attempted to look at an intermediary account (Clarke's Reasonable Expectations View); however, again, little progress was made towards our ultimate goals and central questions and some cases still posed problems for Clarke's intermediary view.

However, perhaps a way to look at the exploration herein, is to take it for what it is and to use the critical examinations as grounds to help further specify a refined set of unique challenges in answering the questions we concerned ourselves with. If nothing else, it may be that we have enumerated and collected various thorny theoretical complications, which any further account of moral responsibility and blame/praise attribution must address. In a sense, the inter-theoretical problems we've explored have condensed and hopefully contextualized something primary to the various accounts, that is, perhaps this work's exploratory nature and ultimately critical conclusions can be taken as a distillation of what (at least a small corner of) the existing literature has already found out. Construed as such, this work can be seen as having furthered the debate, if only humbly locating and re-presenting its specific host of

challenges and occasionally refining/re-contextualizing some of the literature's most problematic cases.

So looking back, which specific challenges, or pressing challenges seemingly need to be factored into any potential further accounts of moral responsibility and praise/blame attributions:

- 1.) *There must be consideration given to how an agent's cognitive features, abilities, and/or character traits came to be configured in the way they exist occurrently (at the time of evaluation).*

I think it would be appropriate to call this the "*Bonnie condition*" of moral responsibility, since if something like an illness brings about the objectionable features of an agent, it is naturally blameless that they come to be the type of person, with the given symptomatic set of features. If a blameless condition can affect one's ability to come to the right moral commitments at a fundamental level (as with Bonnie) then any account of blame/praise and moral responsibility needs something further than to accurately establish the existence of an agent's morally objectionable commitments, simpliciter. The prior question is: how did the commitments come about to begin with? What conditions affected one's ability to arrive at these commitments? Etc. The Bonnie condition is a necessary pre-requisite to morally evaluating an agent, despite what some Q of W theorist may be tempted to contend to the contrary.

2.) *Moral evaluations (both responsibility and blame/praise attribution) must not be solely circumscribed to an agent's own rational (clear-eyed) deliberations about what it was right for the agent themselves to do.*

Perhaps this might be called the "*Finn*" condition of moral evaluation, since it unambiguously comes from Arpaly and Schroeder's case examination of Mark Twain's famous fictional character. If "*inverse akrasia*" can be the subject of praise, then our second stipulation seems more than rightly held. Any account seemingly needs a framework of moral evaluation that can accommodate analysis of irrational mental states and unconscious psychological dispositions. Finn's aversions to acting as he thought to be right, is praiseworthy and we can't have a complete account moral evaluation if we restrict it to only those things which agent's rationally and self-reflectively decide upon as being all-things-considered wrong/or right.

These collections of inter-theoretically pertinent conditions may be the true fruit of this work, even if that is a somewhat unsatisfactory and limited contribution. It hopefully elucidates something primary to all accounts and speaks to issues which escape the domain of any particular view. If these conditions are unsatisfactory or obvious, perhaps the work has at least given a short list of test cases for those looking to provide further positive accounts. I'd say with certainty that if one's account can respond to the combined cases of *Bonnie and Finn*, then it is at least a promising candidate.



## Works Cited:

- Arpaly, Nomy et Schroeder, Timothy. "Praise, Blame and the Whole Self." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 93, No. 2 (Feb., 1999), pp. 161-188.
- Robinchaud, Philip et Wieland, Jan Willem. Eds, *Responsibility: The Epistemic Condition*, First ed., Oxford University Press, 2017.
- Clarke, Randolph. "Ignorance, Revision, and Commonsense." Robinchaud et Wieland, pp. 233-251.
- Talbert, Matthew. "Akrasia, Awareness, and Blameworthiness." Robinchaud et Wieland, pp. 47-64.
- Wieland, Jan Willem. "Introduction: The Epistemic Condition." Robinchaud et Wieland, pp. 1-28.
- Zimmerman, Michael J. "Moral Responsibility and Quality of Will." Robinchaud et Wieland, pp. 219-232.
- Rosen, Gideon. "Culpability and Ignorance." *Proceedings of the Aristotelian Society, New Series*, Vol. 103 (2003), pp. 61-84.
- Smith, Angela M., "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics*, Vol. 115, No. 2, January 2005, pp. 236-271.
- Smith, Holly. "Culpable Ignorance." *The Philosophical Review*, Vol. 92, No. 4 (Oct., 1983), pp. 543-571.