Holly M. Smith

Non-Tracing Cases of Culpable Ignorance

Philosophy Department
Rutgers University
1 Seminary Place
New Brunswick, New Jersey 08901

hsmith@philosophy.rutgers.edu
Tel: 732-932-1321
Fax: 732-932-8617

Abstract

Recent writers on negligence and culpable ignorance have argued that there are two kinds of culpable ignorance: tracing cases, in which the agent’s ignorance traces back to some culpable act or omission of hers in the past that led to the current act, which therefore arguably inherits the culpability of that earlier failure; and non-tracing cases, in which there is no such earlier failure, so the agent’s current state of ignorance must be culpable in its own right. An unusual but intriguing justification for blaming agents in non-tracing cases is provided by Attributionism, which holds that we are as blameworthy for our non-voluntary emotional reactions, spontaneous attitudes, and the ensuing patterns of awareness as we are for our voluntary actions. The Attributionist explanation for why some non-tracing cases involve culpability is an appealing one, even though it has limited scope. After providing a deeper account of why we should take the Attributionist position seriously, I use recent psychological research to argue for a new account of the conditions under which agents are culpable for straightforward instances of blameworthy acts. That account is extended to blameworthiness for non-voluntary responses. I conclude that even when the agent’s failure to notice arises from a non-voluntary objectionable attitude, very few such cases are ones in which Attributionism implies that the agent is blameworthy for her act.

Keywords

Attributionism, blame, blameworthiness, culpability, culpable ignorance, defenses, excuses, ignorance of fact, moral responsibility, negligence, non-tracing, Volitionism
Non-Tracing Cases of Culpable Ignorance

“We are complex beings; no one is all good or all bad.”
- Van Jones (2010)

A. Culpability in non-tracing cases

Recent writers on culpable ignorance have argued that acts done from culpable ignorance can be divided into two kinds: tracing cases, in which the agent’s ignorance traces back to some culpable act or omission of hers in the past that led to the current act, which therefore arguably inherits the culpability of that earlier failure; and non-tracing cases, in which there is no such earlier failure, so the agent’s current state of ignorance must be culpable in its own right. In the law, a negligent act is an act in which the agent fails to advert to some risk of harm and so acts in a way that produces or could produce this harm. Arguably such acts may involve either tracing or non-tracing culpability.

My chief focus in this paper is the question of moral culpability in non-tracing cases. Whether or not there are non-tracing cases in which the agent is genuinely culpable for her ignorance, and therefore her act, is a matter of some controversy. Here are some cases that have been offered as examples of non-tracing cases in which it is claimed we would normally judge that the agent is responsible, and blameworthy, for what she does.

*Hot Dog.* Alessandra, a soccer mom, has gone to pick up her children at their elementary school. As usual, Alessandra is accompanied by the family’s border collie, Bathsheba, who rides in the back of the van. Although it is very hot, the pick-up has never taken long, so Alessandra leaves Sheba in the van while she goes to gather her children. This time, however, Alessandra is greeted by a tangled tale of misbehavior, ill-considered punishment, and administrative bungling which requires several hours of indignant sorting out. During that time, Sheba languishes, forgotten, in the locked car. When Alessandra and her children finally make it to the parking lot, they find Sheba dead from heat prostration.

*On the Rocks.* Julian, a ferry pilot, is nearing the end of a forty-minute trip that he has made hundreds of times before. The only challenge in this segment of the trip is to avoid some submerged rocks that jut out irregularly from the mainland. However, just because the trip is so routine, Julian’s thoughts have wandered to the previous evening’s pleasant romantic encounter. Too late, he realizes that he no longer has time to maneuver the ferry.

---

1 I earlier dealt with tracing cases in Smith (1983).

2 *Hot Dog* and *On the Rocks* are taken from Sher (2009), p. 24. In *Hot Dog* I have changed the upshot for the dog from “unconscious from heat prostration” to “death” for later ease of description.
Loaded Gun. Fifteen year old Luke’s father keeps a shotgun in the house for hunting, and last fall started to teach Luke how to shoot with it. Luke is proud of the gun and takes it out to show his friend. Since the hunting season has been over for months, it doesn’t occur to Luke that the gun might be loaded. In showing the gun to his friend, unfortunately he pulls the trigger while the gun is aimed at his friend’s foot, blowing the friend’s foot off.3

Forgotten Birthday. Angela forgot a close friend’s birthday. A few days after the fact, she realized that this important date had come and gone without her so much as sending a card or giving her a call. Angela was mortified. What kind of friend could forget such a thing? Within minutes Angela was on the phone to the friend, acknowledging her fault and making apologies (A. Smith 2005, p. 236).

Driveway Disaster. Nate, tired from waking up early, backs out of his driveway. His thoughts turn to his meetings that day, and his attention is partly focused on a radio commercial. Due to his inattention, Nate doesn’t see a child walking to school and so hits him, breaking the child’s leg (King 2009, p. 578).

Of course, in each of these cases one can debate whether it is really a non-tracing case. For example, in Hot Dog one could claim that Alessandra was culpable for bringing the dog with her on such a hot day, knowing that she usually leaves the dog in the car. Alternatively one could claim that she was culpable for not bringing the dog into the school once she arrived. Her culpability for the dog’s death, then, would trace back to her culpability for bringing the dog in the first place, or for not bringing it into the school. In Loaded Gun one could argue that Luke was culpable for pulling the trigger because he was reckless (not merely negligent) in acting without knowing for certain whether the gun was loaded.4 Or one could claim that he was culpable for not checking whether the gun was loaded before pulling the trigger, so that his culpability for injuring his friend’s foot would trace back to his culpability for not checking whether the gun was loaded. In Forgotten Birthday one could claim that Angela was culpable for not writing her friend’s birthday on the calendar to remind her of the important date. Her culpability for forgetting the birthday, then, would trace back to her culpability for not making a notation on her calendar. However, such claims are less plausible in On the Rocks and Driveway Disaster, partly since Julian and Nate have no reason to believe they will be negligent in piloting the boat or backing out of the driveway, given that this has never happened to them before. In light of their past good records, perhaps they have no reason to take special precautions. Moreover, it could be argued that even if Alessandra is culpable for bringing the dog with her, her culpability for this is less severe than her culpability for forgetting that the dog is in the car (since the known probability, on setting out, that she will forget the dog in the car – since she could always decide to take it into the school -- is less than the probability, on entering the school, that she will forget it

---

3 This is a variant on a case described by Sher (2009), p. 7.
4 On the common legal definition one acts recklessly when one acts in conscious disregard of a substantial and unjustifiable risk that one’s conduct will harm another. See The American Law Institute Model Penal Code (1985), Section 2.02. If Luke was conscious of the risk that the gun was loaded, his action is reckless.
once inside). If so, her culpability for bringing the dog with her cannot fully explain her greater culpability for forgetting the dog when she’s inside the school. Perhaps the full extent of her blameworthiness for forgetting the dog can only be fully explained by ascribing some non-tracing culpability to her.⁵ Reflecting on such cases leaves me uncertain whether genuine non-tracing cases really occur. However, for the sake of argument, let us concede, at least provisionally, the existence of non-tracing cases in which (a) an agent acts badly because she failed to notice some important aspect of her prospective action, and (b) her blameworthiness cannot be fully explained by any past culpable act or omission on her part.

These non-tracing failure-to-notice cases, as I shall call them, are highly puzzling, since they violate a standard conception of what is required for an agent to be blameworthy for her action. According to this standard model, an agent is only blameworthy for an action or its consequences when either she was consciously aware at the time she acted that the action would be wrong, or was consciously aware that in acting she unjustifiably risked doing something wrong.⁶ In the failure-to-notice cases there is no such awareness. How, then, can the agent be culpable for her action? Why doesn’t her lack of awareness excuse her, as it does in cases of non-culpable ignorance?

One promising emergent strategy for defending the claim that the agent is indeed culpable in a non-tracing failure-to-notice case is to base the claim on the views of a group of contemporary writers (often called “Attributionists”) who argue that we are often as responsible, or blameworthy, for our non-voluntary emotional reactions, spontaneous attitudes, and patterns of awareness as we are for our voluntary actions.⁷ Attributionists hold that someone who reacts with dismay on discovering that the dentist to whom she had been recommended is African-American is blameworthy for her racist emotional response. She is blameworthy for her emotional response in the same way that someone is blameworthy who performs a bigoted act in failing, for racist reasons, to hire the better-qualified candidate as secretary. Attributionists argue that there are many cases in which a person’s feelings are not merely bad but also blameworthy, and that the person is to blame for having those feelings.⁸ Their argument rests heavily on examples of cases in which we do seem to make judgments of blameworthiness (both first-person and third-person) for such mental occurrences, for example holding that Angela is blameworthy for forgetting her friend’s birthday, and that the dental patient is blameworthy for her bigoted dismay at the race of her new dentist. Of course, we have many mental states and

---

⁵ Moreover, as Sher suggests, understanding *Hot Dog* as a tracing case only pushes back the perhaps equally-hard question of whether Alessandra was culpable for not considering whether to bring the dog with her when she set out. Sher (2009), p. 83.
⁶ For a vivid description of this view, which he calls the “Searchlight View,” see Sher (2009), Chapter 1. See also discussion in Moore and Hurd (DATE).
⁷ The term “Attributionist” appears to trace back to Gary Watson’s (2004; originally published 1996) discussion of two different forms of blame, ‘aretaic blame’ (in which we attribute or impute conduct to an individual) and ‘accountability blame.’ Proponents of Attributionism include A. Smith (2005) and other pieces; Hieronymi (2008) and other pieces; Adams (1985); Frankfurt (1988); Williams (1973); and Scanlon (1998). There are important differences among these philosophers; I shall be guided primarily by Angela Smith’s account.
⁸ See, for example, the debate between Levy (2005) and A. Smith (2005), and Smith’s (2008) response to Levy.
occurrences that have little or no moral significance. The fact that you feel pleased when you hear your favorite song on the radio reflects nothing morally significant about you. However, Attributionists focus on the set of non-voluntary mental states and activities that reflect morally significant “evaluative judgments” held by the agent: bigoted or tolerant attitudes, valuing or disvaluing of friends’ well-being, concern or lack of concern over individuals’ needs or discomfort, amusement at ethnically insulting jokes, pleasure or pain in the success of others, discounting the minor moral flaws of others, admiration or resentment of the good qualities of others, and so forth. According to Attributionists, these evaluative judgments form our “moral personality” and therefore can be held to our credit or discredit. The non-voluntary responses that reflect these evaluative judgments are ones over which we typically have little control. Precisely for this reason, they often provide a better window into our souls than many actions, which can be strategically performed in order to disguise our real feelings. On the Attributionist view, if a person is blameworthy for an evaluative judgment, then the person is also blameworthy for any non-voluntary reaction or response that arises from this evaluative judgment.

If the Attributionists are correct in saying that we can be blameworthy for what, following Angela Smith, I shall call our non-voluntary responses, then there is clearly an Attributionist argument that an agent can be blameworthy for her non-tracing non-voluntary failure to notice something morally significant (Scanlon 1998, pp. 268-9). In *Hot Dog*, for example, Alessandra doesn’t remember the dog in the hot car. Suppose her failure to think of the dog’s plight cannot be traced to some earlier dereliction, and instead reflects an objectionable evaluative attitude, such as insufficient concern for the dog’s well-being. Then, on the Attributionist view, her failure to think of the dog’s plight is blameworthy because it reflects this underlying blameworthy evaluative attitude. By extension her leaving the dog in the car too long is a blameworthy act because it arises from her blameworthy failure to think of the dog’s plight. On this account her “ignorance” is culpable and so makes her act culpable as well. The act is culpable, and

---

9 Of course we often make evaluative judgments that do not reflect at all on our moral character; my lingering over the thought of a chocolate soda, and not lingering over the thought of a strawberry soda, shows I prefer chocolate to strawberry, but this reveals nothing about my moral preferences or values. Attributionists should only be interested, for purposes of assigning moral responsibility and blame, in evaluative attitudes that reflect moral attitudes.

Angela Smith emphasizes that these evaluative judgments may not be conscious, or be consciously recognized by the person who holds them, who may only discover she has such values through her own non-voluntary responses to events. See Smith (2005), pp. 251-52. Smith adds another qualification on which mental states can count as ones for which a person can be blameworthy. She says “In order for a creature to be responsible for an attitude, on the rational relations view, it must be the kind of state that is open, in principle, to revision or modification through that creature’s own processes of rational reflection” (p. 256). (See also Scanlon (1998), pp. 20-24, 272.) For an Attributionist like Smith, who opposes theorists holding that a person is only responsible for those attitudes that are in some way connected to her choices (either her past choice to have this attitude, or her current choice to endorse it, or her ability to modify it in the future), this restriction brings her uncomfortably close to insisting on the very kind of voluntarism that she rejects among her opponents. For this reason I shall not emphasize it.

10 The term “moral personality” is used by Hieronymi (2008), p. 361.

11 Note that “non-voluntary,” rather than “involuntary” seems to be the correct term here: typically our emotional reactions, etc., are not contrary to our will, which is what “involuntary” may suggest. See Hieronymi (2008), p. 368.
for the same reasons, as it would be if it had arisen from a conscious decision on her part to leave the dog in the hot car because she doesn’t value its welfare very highly. In both cases the harmful act arises, ultimately, from a reprehensible evaluative attitude on the agent’s part. The only difference is that in the failure-to-notice version of Hot Dog a culpable state of “ignorance” mediates between the reprehensible evaluative judgment and the risky act.

In criticizing the Attributionist view, Matt King points out that it is often simply not credible that the agent in a non-tracing failure-to-notice case has the alleged underlying objectionable evaluative attitude (King 2009, pp. 583-6). Alessandra, for example, might have a long history of displaying great love and concern for the family dog, spending hours playing with her, training her, obtaining expensive veterinary care when the dog is ill, and so forth. Her appalled response on discovering the dog dead from heat prostration further manifests her high value for the dog’s welfare, and the improbability that she was even willing to risk exposing the dog to danger. There could be similar information about Julian, Luke, Angela, and Nate, in each case providing overwhelming evidence that the agent did not hold any objectionable evaluative judgment that might be inferred from his or her failure to notice or remember. In such cases, no stable faulty attitude could be attributed to the agent in light of his or her one-time failure to take notice. Indeed it may not even be plausible to ascribe to the agent a momentary faulty attitude of the kind shown by an exhausted soldier who shoots at a movement in the house he is searching, too tired to care about the risk that he is shooting an innocent civilian rather than an enemy combatant.

In cases such as these, in which we can’t reasonably impute a faulty evaluative attitude to the negligent agent, the Attributionist strategy for imputing blame to the agent for her culpably ignorant act seems to fail. However, it may succeed in other cases, in which we can reasonably infer that the failure-to-notice does derive from an underlying objectionable evaluative attitude. For example, consider another case:

**Bad Joke:** Ryland is very self-absorbed. Though not malicious, she is oblivious to the impact that her behavior will have on others. Consequently, she is bewildered and a bit hurt when her rambling anecdote about a childless couple, a handicapped person, and a financial failure is not well received by an audience.

---

12 Typically Attributionists don’t address the question of whether the culpably ignorant agent is as culpable as the witting agent who performs the same act.

13 Note that on this account the agent’s ignorance has no intrinsic moral status; its blameworthiness is derived from the evaluative attitude that generates it. For this reason one can see such “non-tracing” cases as more closely aligned with “tracing” cases than it appears at first blush; in both cases the moral status of the ignorance traces to something else (a motivational state) that has intrinsic moral status.

14 Sher himself concedes that Alessandra may have no morally objectionable attitude towards the dog (Sher 2009, p. 131).

15 Inference to a low concern for the victim’s well-being is even less plausible in the horrific cases of parents who have forgotten their infant in the back seat of the car and returned at the end of the day to discover the child dead from heat exposure. For accounts of some of these cases, see Weingarten (2009).
that includes a childless couple, a handicapped person, and a financial failure (Sher 2009, p. 28).

From Sher’s description it is clear that Ryland has a standing tendency not to notice the responses of people around her because she is so focused on herself. In this case the Attributionist’s analysis of the case is more plausible, because Ryland’s failure to notice does derive from an objectionable valuation of the importance of her own as compared to other people’s welfare.16 But the fact that we have to distinguish the earlier cases from Ryland’s case shows that the Attributionist strategy for explaining culpable ignorance in non-tracing cases is at best severely limited. The cases we began with – the cases of Alessandra, Julian, Luke, Angela, and Nate – are ones in which many people, including the typical Attributionist, feel the agents are indeed blameworthy. The fact that the Attributionist cannot explain why the agents in those cases (who cannot plausibly be said to have reprehensible17 evaluative attitudes) are blameworthy shows that the Attributionist account of what makes the agents culpable in failure-to-note cases doesn’t account for all the intuitions that apparently need to be captured. Nor does it appear that Attributionism has any resources for explaining why we may feel that the agents in these original cases are blameworthy. Even if some independent explanation can be found, the prospects look bleak for achieving anything like a unified account of what makes agents blameworthy in non-tracing failure-to-notice cases.18

Despite its limited scope, the Attributionist explanation for why some non-tracing cases involve culpability is an appealing one. In this paper I will examine this

---

16 Note, however, that it is also plausible to see this case as a tracing case, rather than a non-tracing case. If Ryland has a standing over-valuation of her own interests as compared with those of other people, this is a moral flaw that she has presumably had ample opportunity to discover and correct in the past. Her failure to do so may well be culpable, so that her culpability for her present personality trait traces at least in part to past derelictions.

17 Throughout this paper I will often use “reprehensible” as a generic term for attitudes that are morally objectionable. The degree to which an attitude is objectionable varies from attitude to attitude, and in normal discourse we tend to reserve the term “reprehensible” for those attitudes that are highly egregious. One can be blameworthy even if one’s attitude is not highly egregious. However, for brevity, I shall use the term “reprehensible” to cover all cases of attitudes that are morally faulty.

18 Sher proposes an account according to which an agent is responsible for an act done in culpable ignorance if the agent “is unaware that the act is wrong or foolish despite having evidence for its wrongness or foolishness his failure to recognize which (a) falls below some applicable standard, and (b) is caused by the interaction of some combination of his constitutive attitudes, dispositions, and traits.” (Sher 2009, p. 88.) However, this account does not seem helpful. It is unclear how we are to determine whether a person’s failure to notice falls below the “applicable standard” without knowing whether the person’s constitutive attitudes, dispositions, and traits fall short of some related standard. After all, the standard cannot simply say “There was evidence that she ought to have taken into account,” since there are many cases in which an individual ought to have taken evidence into account but is not culpable for failing to do so. But Sher seems to deny that the person’s character (or even a component of that character) must be morally flawed for her to be responsible – it must simply be the case that her underlying character gives rise to the failure to notice (pp. 86-91). Thus we cannot invoke any moral flaw in the person’s constituent character to explain why she falls short of the standard. Taken as a whole, this account gives us no clear handle on the difference between people who are blameworthy and those who are not for their failures to notice.
explanation more thoroughly, aiming to make out the best possible case for the Attributionist approach. My first task will be to provide a deeper argument for why we should take the Attributionist position seriously. Taking into account recent discoveries in social psychology, I will then provide a new and more sophisticated account of the conditions under which agents are culpable for straightforward instances of blameworthy acts. That account will be extended to cases in which the agent’s culpability is for a non-voluntary response rather than an action. Finally, I will draw lessons from this for putative cases of non-tracing culpable ignorance. My conclusion will be that, even among the cases in which the agent’s failure to notice arises from an objectionable attitude, very few such cases are ones in which the agent is blameworthy for her act.

B. Filling out the Attributionist account of culpability in non-tracing cases

Let us examine the Attributionist account more closely to see how well it holds up under scrutiny.

Of course Attributionists are in the minority in holding that people can be blameworthy for non-voluntary responses. They are opposed by “Volitionists,” who claim that we can only be responsible or blameworthy for events over which we have voluntary control, such as our acts and their foreseen consequences. Volitionists concede that sometimes we do have indirect control over our emotions and patterns of awareness (for example, an airplane pilot may have trained herself to be constantly alert to the plane’s airspeed indicator, or an emergency room physician may, through long exposure, overcome her natural feeling of horror at the sight of blood). In such cases Volitionists would agree that we are responsible for any objectionable mental states that we have knowingly brought about in this indirect fashion, and they would agree that we should train ourselves to have beneficial patterns of awareness, or to eliminate harmful ones. However, they would argue we have no direct control over emotions or patterns of awareness at the time they occur, and for the most part we have no indirect control either. Typically we haven’t chosen to create, maintain, or eliminate these parts of our mental life: they simply occur, and in some cases they are beyond any control we may wish (or may have wished) to exert over them. In cases in which the person has neither direct nor indirect control over an emotional response or a failure of awareness, Volitionists deny that the person has any responsibility, and certainly deny that he is blameworthy, for these occurrences (Wallace 1996, pp. 131-132, Sidgwick 1907, pp. 59-61; and Taylor 1970, pp. 241-52). Volitionists can, and sometimes do, grant that we can evaluate a person as bad insofar as she has a morally objectionable emotional response or failure of awareness, just as we can evaluate a dog as bad insofar as it has a nasty temper. But dogs are not blameworthy for their nasty tempers, and the Volitionist claims that usually people are not blameworthy for their emotional responses. Volitionists insist that we make a variety of kinds of moral evaluations of people, and that evaluating someone (or

---

19 I take these labels from Levy (2005). Although I will confine my attention to questions about blameworthiness, most discussions of the contrast between Attributionism and Volitionism note that moral responsibility extends beyond blameworthy acts to praiseworthy and morally neutral acts (and emotions), and therefore encompasses acts (or emotions) that are neither blameworthy nor praiseworthy.
her feelings, or her patterns of awareness) as bad or contemptible is quite different from evaluating her (or her mental states) as blameworthy.

The Volitionist argument that we cannot be blameworthy for our non-voluntary responses appears to be a persuasive one. What counter-argument can the Attributionist provide for holding that individuals are blameworthy for their non-voluntary emotional and cognitive responses, as well as for their acts? As I noted before, Attributionists themselves seem to rely mainly on examples to elicit intuitions supporting their position. But it is possible to propose a more principled argument for their position. This argument is based on the observation that there is a deep parallel between two common psychological processes. On the one hand, there is the process involving an agent’s decision to act and the action this decision gives rise to. On the other hand, there is the process involving an agent’s evaluative attitude and the non-voluntary response the attitude gives rise to. In the first case the agent’s mental state (perhaps a combination of beliefs and desires, or perhaps the formation of an intention) causes an act (such as raising a hand, or lifting a suitcase). In the second case the agent’s mental state (an evaluative attitude) causes either another mental state (for example, a failure to notice, or a rush of dismay), or a non-voluntary bodily state (for example, a response such as blushing or gasping, a facial expression, or a feeling of revulsion). It may strike one that there is a strong parallel between these two kinds of cases, and it may be tempting to say that if the first kind of process (which we can call the “desireÆact process”) leads to an act for which the agent is morally responsible, then the second kind of process (which we can call the “attitudeÆresponse process”) can similarly lead to a response for which the agent is responsible. It is sometimes said that an agent is the “author” of her actions, and so is responsible for them – but it is tempting to say that an agent is equally the “author” of her non-voluntary responses, and so is responsible for them. In both cases morally significant mental states lead to the event in question.

Of course the Volitionist will deny that the parallel between the two processes is strong enough to support the conclusion that individuals are responsible for non-voluntary cognitive and emotional responses as well as for their acts. But what significant differences between the act-producing process and the response-producing

---

20 Attributionists emphasize that in some (but not all) of the cases in which the agent is responsible for her non-voluntary response, the response may be partly constituted by the judgment. For example, the emotion of jealousy is not only caused by a certain judgment, it involves the judgment, and would not count as jealousy unless the agent made that judgment. (See, for example, A. Smith (2005), p. 258.) However, Attributionists implicitly admit cases in which the non-voluntary response is merely caused by the judgment (such as failures-to-notice cases) rather than constituted in part by it, although they distinguish these from other bodily reactions that are merely caused by judgments, as one’s nausea before a public speaking performance is caused by one’s fear of public speaking. See Smith, p. 257-59. I think the discussion of the difference between these kinds of cases is as yet inadequate. In any event, these theorists’ discussions tend to render somewhat fuzzy the distinction between the initiating mental state (the evaluative judgment) and the non-voluntary response it gives rise to. I shall have to leave this distinction somewhat fuzzy as well; the important point is that the Attributionists regard the agent as blameworthy for the initiating evaluative judgment as well as for any distinct non-voluntary response to which it gives rise.

21 See, for example, Gary Watson (2004), p. 263.
process can the Volitionist call on to support her claim that the first kind of process involves a voluntary event and can be blameworthy, while the second kind of process does not involve a voluntary event and cannot be blameworthy?

In defense of her position the Volitionist might say that the first but not the second kind of process results in a blameworthy occurrence because, while the agent can resist the process by which a desire leads to the performance of the desired act, no agent can resist the process by which her underlying evaluative attitudes give rise to a non-voluntary response such as blushing. However, it’s not clear how persuasive this argument is. Not all desires to perform an act can be successfully resisted (remember how many dieters break their diets). Even when desires can be resisted, this may only mean that a second desire (for example, a desire to stay thin) turns out to be stronger than the first desire. Or it may mean that a second-order desire (not to act on one’s first desire) is stronger than the first desire. It hardly follows that one can always resist a desire to act, or that one can resist one’s entire constellation of desires. And, on the other hand, people can and often do control their non-voluntary reactions: squelching a grasp, restraining a scream under torture, maintaining a poker face during stressful situations, or suppressing a wince when receiving a flu shot. The appeal to “capacity to resist” does not appear to do the work needed for the Volitionist.

Second, the Volitionist might claim that a desire or intention to act is itself voluntary, and so transmits its voluntary status to the act it gives rise to, whereas evaluative attitudes are non-voluntary, and so have no such status to transmit. However, the Volitionist is on highly tenuous grounds here. Many people believe that desires and intentions are not typically themselves voluntary, but ultimately arise from our biology, our environments, and our prior experiences. Moreover, some Attributionists, quite correctly, explicitly include desires and intentions among the typically non-voluntary states that can give rise to further non-voluntary responses. And in any event it is clear that many of the “evaluative attitudes” to which the Attributionist appeals -- greed, fear, envy, admiration, resentment, and so forth -- are the very kinds of motivational states that can give rise to actions as well as to non-voluntary responses. People often act out of fear or greed. So if these states are voluntary, this status would be transmitted to so-called “non-voluntary” responses as well as to actions; on the other hand, if they are not voluntary, they have no status to transmit to either actions or non-voluntary responses. Hence this second argument does not seem to help the Volitionist’s case.

Third, the Volitionist might claim that rationality is an important aspect of blameworthiness, and claim further that intentions, but not evaluative attitudes, partake of rationality. The claim would be that intentions have a rational aspect in the sense that they can be changed by the agent’s acquiring new information: no rational person, intending to put sugar in her coffee, would maintain that intention on discovering that what she thought to be sugar is actually arsenic. It is certainly true that intentions are sensitive to information in this manner. But so, often, are non-voluntary evaluative

\[\text{22 For an accessible account of contemporary psychologists’ work on “emotional regulation,” see Carey (2010).}\]

\[\text{23 See, for example, Hieronymi (2008), p. 368.}\]
attitudes: one’s admiration for a colleague may dissipate on learning that she has been defrauding the company, and one’s loyalty to a friend may increase on learning that he has been quietly working to support one’s career advancement. Of course such evaluative attitudes are derived from one’s basic evaluative attitudes together with relevant information (perhaps one has a basic admiration for honesty, and so derived one’s initial admiration for one’s colleague from evidence that she is honest.) And basic evaluative attitudes may be harder to change through acquisition of information (in this case one still values honesty, but simply corrects one’s earlier derived admiration for one’s colleague). Nonetheless one can sometimes change these basic attitudes also, as when someone comes to be convinced about the moral importance of animals’ pain or the life of a human fetus. What this discussion suggests is that in considering sensitivity to new information we should distinguish basic from derived intentions, and basic from derived evaluative attitudes. Having made these distinctions we can recognize that both derived intentions and derived evaluative attitudes are highly sensitive to changes in information, while basic intentions and basic evaluative attitudes are less readily altered by exposure to new information, although they can sometimes be altered nonetheless. However, there seems to be nothing in the realm of sensitivity to information that points to an important distinction between the “rationality” of intentions and the “rationality” of evaluative attitudes, or that implies intentions can provide a foundation for blameworthiness while evaluative attitudes cannot.24 The Volitionist’s third argument falls short as well.

The Volitionist’s claim that there is a significant distinction between the process by which desires or intentions lead to actions, on the one hand, and the process by which evaluative attitudes lead to non-voluntary responses, on the other, is seriously undermined by the fact that desires and intentions can lead, not just to actions, but also to non-voluntary responses. Thus my desire to eat the last donut may lead to my reaching for the donut. But it may also lead to non-voluntary responses: my continuing to think about the donut, my salivating at the thought of eating the donut, and my feeling cross if someone else takes it before I do. My intention to euthanize my sick cat may lead to my taking the cat to the vet, but may also lead to my blood pressure’s rising as I tell the vet my decision. Moreover, as we saw before, it is equally clear that my “evaluative attitudes” may lead, not just to non-voluntary responses, but also to acts: my feeling of superiority to an acquaintance may lead to my adopting a sneering expression when I talk about her, but it may also lead to my not inviting her to a social event. Thus the initiating elements in both kinds of processes are capable of resulting in both kinds of upshots, acts and non-voluntary response.25 This renders it more difficult to discern a morally significant difference between the two kinds of process.

24 As we have seen, Attributionists often stress the fact that the evaluative attitudes for which they claim you are responsible are attitudes that are “open, in principle, to revision or modification through that creature’s own processes of rational reflection” (A. Smith 2005, p. 256).
25 It is worthwhile noting a further point in the comparison between the act-producing process and the response-producing process. We can often control the way we perform a basic act – for example, I can control whether I raise my arm quickly or slowly. But it is also true that we can sometimes control the way we express an evaluative attitude. I can control whether I grimace in an exaggerated or restrained manner, and can sometimes modulate my look of shock or disgust. These kinds of cases may show that the boundary between acts and non-voluntary responses can
There is, of course, a prominent and genuine difference between the two kinds of processes. In the intention→act process, the initiating psychological state (the desire or intention) involves the aim to do something. In the typical case this aim leads to an effect (the act) that accurately fulfills the aim. Thus you intend to raise your arm, and you raise it, thereby fulfilling your aim. You intend to turn on the light, and you turn it on. Your intention causes the intended act. Of course this sometimes misfires. There are cases in which, because of physiological problems or misinformation, your intention does not issue in any act, or in the precise act that you intended. Nonetheless, in the paradigmatic case, the intention causes what it aims at, at least at the level of the basic act.\(^{26}\) However, the evaluative attitude→non-voluntary response process does not follow this pattern. The initiating psychological state does not necessarily have an aim, much less an aim that is fulfilled by the resultant response. You feel embarrassed, and this causes you to blush. Your feeling of resentment gives rise to a scowl. You feel superior, and so your face assumes a sneer. You admire someone, and so you feel a special gratitude when that person praises you. Although your initiating evaluative attitude has propositional content (for example, the proposition that you are socially superior to someone), it need not involve anything like an aim. Embarrassment doesn’t involve the aim of blushing, even though it gives rise to blushing. Resentment doesn’t involve the aim of scowling, even though it gives rise to scowling. We have seen that a desire can serve as the initiating psychological state for either the intention→act process or the evaluative attitude→non-voluntary response process. But in such cases, where the initiating evaluative attitude does involve an aim (for example, greed may involve the aim of securing the last donut, and may give rise to the act of taking the donut), the non-voluntary response that it gives rise to (for example, salivating, or thinking obsessively about the donut) typically does not itself involve satisfying that aim.\(^{27}\) Thus there is an important distinction between the types of mental states that initiate the intention→act process (they must involve aims to bring something about) and those that initiate the evaluative attitude→non-voluntary response process (they need not involve an aim to bring anything about). And there is a further, even more important distinction between the types of events that are the upshots of the intention→act process (they paradigmatically involve satisfying the initiating state), and the types of events that are the upshots of the evaluative attitude→non-voluntary response process (paradigmatically they do not satisfy any aim of the initiating state). These distinguishing features of the intention→act process are an important part of what make it appropriate to refer to the resulting act as “voluntary,” since the evaluation.

---

26 Note that an aim to “bring something about” can include an aim to omit doing something. For an account of the distinction between “basic” and “higher level” acts, see Goldman (1970).

27 Note that, as pointed out before, an evaluative attitude such as greed may give rise both to a non-voluntary response (which doesn’t necessarily satisfy it) and to an action (acquiring the desired objects) which paradigmatically does satisfy it.
voluntary is that which proceeds from, and carries out, the will. These features are absent from the evaluative attitude non-voluntary response process, so it is inappropriate to refer to the resulting response as “voluntary.”

From this we see why it is appropriate to view the emotional and other responses to evaluative attitudes as distinctively non-voluntary, even in cases where the agent manages to exert some degree of control over them by, for example, suppressing his impulse to scream or scowl. Traditionally theorists have attempted to capitalize on this way of understanding “voluntary” to show that, because non-voluntary responses aren’t voluntary in this sense, they cannot be blameworthy. However, in the context of our discussion, I don’t immediately see how to do so in a non-question-begging way. Unless you take “blameworthy” to entail “voluntary” it doesn’t seem to follow directly from the fact that these responses are non-voluntary (something admitted up front by Attributionists) that the agent is not responsible for them and not blameworthy if the underlying attitude is morally objectionable. Of course the Attributionist denies any such entailment. Hence merely from this distinction between the act-producing process and the response-producing process I cannot see a straightforward argument that the non-voluntary product of the latter is something for which the agent cannot be blameworthy.

This discussion suggests that many kinds of considerations that might be adduced by the Volitionist to justify drawing a strong distinction between the act-producing process and the response-producing process -- a distinction that would justify viewing the upshot of the act-producing but not the upshot of the response-producing process as blameworthy -- do not seem to provide a foundation for the desired conclusion. The two processes are much more similar than we might think at first blush. And even the respect in which they clearly differ, which makes it appropriate to classify the act-producing process but not the response-producing process as “voluntary,” does not, without begging the question against the Attributionist, seem to establish that the two processes differ with respect to their role in establishing blameworthiness. The Attributionist can capitalize on these similarities to reject the Volitionist’s argument that we cannot be blameworthy for our non-voluntary responses.

C. A closer look at blameworthiness

We have seen that the most natural line of argument that agents are not blameworthy for their non-voluntary responses runs afoul of the manifest similarities between the act-producing process and the response-producing process. However, it doesn’t follow from this that non-voluntary responses can be blameworthy. They may fail to meet other important necessary conditions for blameworthiness. Hence, before we conclude that we can be blameworthy for our non-voluntary responses, and in particular for our failures to notice, we need to look more deeply at what is required for being blameworthy in general. This requires us examine more closely what conditions are necessary for an agent’s blameworthiness in straightforward cases of culpability for voluntary actions.

28 Setting aside “deviant causal chain” cases, and cases in which the will itself is unfree.
29 This is sometimes expressed by positing a “control” condition for moral responsibility.
Let us start with a widely accepted Humean-style account of the circumstances under which an agent is blameworthy for an act. This account may be stated as follows:

I. S is blameworthy for performing act A if, and only if:

1. Act A is objectively wrong,
2. S had a reprehensible configuration of desires and aversions, and
3. This configuration gave rise to the performance of A.30

Thus, for example, a teenager who wants to humiliate a classmate launches a vicious Facebook attack on the classmate. The teenager is blameworthy for launching the attack, because it arose from a reprehensible motive and resulted in the planned objectively wrong act.

Suppose, however, that the teenager composes her attack and hits the “enter” button, but due to a software malfunction the humiliating screed does not get posted on her Facebook page, and no damage is done to the classmate. According to the account just given, the teenager is not blameworthy, because there is no objectively wrong act to which her reprehensible motive gave rise.31 Nonetheless there is still an obvious sense in which we want to hold the teenager to blame, since she knowingly decided to perform a wrongful act. This suggests that we can usefully distinguish between an agent’s being blameworthy for an objectively wrong action, and an agent’s being blameworthy for a choice to perform an action (which may or may not result in the intended action).32

Suitably adjusting the above account, we might propose the following conditions for blameworthiness for such a blameworthy choice:

II. S is blameworthy for the choice to perform act A if, and only if:

1. S chose to perform act A,
2. S had a reprehensible configuration of desires and aversions, and
3. This configuration gave rise to S’s choice to perform A.

On Account II, the teenager is blameworthy for her choice in deciding to launch the attack on her classmate, even if no harm issues from the decision because of the software

30 This account is taken from H. Smith (1983), p. 556. In the original account I used the term “is to blame” rather than “is blameworthy.” As I point out in this earlier work, the process by which the desires and aversions give rise to the act must not involve what has come to be called a “deviant” causal chain. In the following text, I will always assume this further condition is met.

31 For certain more severe types of offenses, such as homicide, the law recognizes not just the offense of carrying out the homicide, but also the offense of attempted homicide. Inspired by this, one might accuse the teenager of performing the objectively morally wrong act of “attempting to humiliate her classmate.” However, it is not clear that morality recognizes such an offense, or needs to do so, so long as it imputes blameworthiness to agents for their bare choices (as I shall argue it should) as well as for their acts.

32 Note that to hold a person is blameworthy is always to negatively appraise the person for some episode, such as an act or choice. Such judgments are distinct from negative evaluations of the person, or the person’s character, as a whole (“Pol Pot was evil”).

15
malfunction. She would be equally blameworthy for the choice if no harm issued from it because of ignorance or misinformation, for example, her believing falsely that hitting the “shift” button would post her screed. Putting together Accounts I and II, we can infer that whenever someone is blameworthy for an act, he or she is also blameworthy for the choice to perform that act. In fact blameworthiness for choices (closely related to what legal theorists sometimes call having a “guilty mind”) is an essential component of blameworthiness for acts.

For much of the remainder of this paper it will be convenient to focus on blameworthiness for choices rather than blameworthiness for acts, so my examples shall involve this notion, without further inquiry into whether or not the choice led to a wrongful action. This is partly because a more careful account of blameworthiness for an action would have to inquire into matters such as the extent to which the nature of the resulting action must match the nature of the action as the agent envisioned it when she made her choice, and such questions are irrelevant for my immediate concerns.

Let us turn, then, to a closer examination of blameworthiness for a choice. This form of blameworthiness depends on the agent’s configuration of desires and aversions being reprehensible. But what is required for a configuration of desires (for brevity I shall omit further mention of aversions) to be reprehensible? One tempting answer, suggested by the teenager’s case, is that the configuration of desires simply consists of a morally bad desire. The desire to humiliate a classmate is a morally bad desire, so we conclude the teenager is blameworthy for her choice to do this.

However, this simple answer is not adequate. Most of our choices, whether for good or ill, are not done from a single desire, but rather arise from a complex configuration of desires, some morally good, some morally bad, and some morally neutral. Of course we might say that a choice counts as blameworthy only if the agent’s configuration of desires included at least one bad motive. However, deeper reflection reveals that sometimes an agent’s choice is blameworthy because she lacks a certain desire, not because of the objectionable nature of any of her actual desires. Thus an instructor is blameworthy for his choice when he, from self-interest, chooses to steal expensive books from a graduate student’s mailbox and sell them online. The instructor’s sole motive of self-interest is not morally bad in itself. What is wrong with him is his lack of any concern for morality or for the damage his act does to the graduate student. His not caring about these matters is a moral flaw, and makes him blameworthy for his choice. Thus the fact that a choice arises at least in part from a bad desire is not necessary for one to be blameworthy for the choice, since it is the lack of certain kinds of desires that makes certain motivational configurations reprehensible.

---

33 I shall not try to say a great deal about what kinds of motivations, desires, or aversions, count as morally good or bad in themselves. A clear-cut example of a morally good motive would be a desire to do what is morally right; but many theorists hold that a desire to perform an action for one of the features that makes it right (such as telling the truth, or assisting a person in need) also counts as a morally good motive. In the text I shall assume this is true.
Just as having a bad desire is not necessary for being blameworthy, so having a bad desire is not sufficient to establish that an agent’s choice is blameworthy. Consider a Coast Guard captain whose duty requires him to rescue survivors of sinking boats. If he cannot rescue all the survivors, his duty requires him to save as many as possible. Here is a case involving such a scenario:

*Coast Guard Rescue I*: Two boats, the *Amelia* and the *Betty*, have capsized in a storm. As he approaches the scene Coast Guard Captain Brian sees that he will have to choose to rescue one group of survivors or the other, and that members of the group he cannot rescue first are doomed to die of exposure, since he will not be able to get to them in time. He estimates that the passengers of the *Amelia* are the larger group, so he heads his cutter in that direction, aiming to save them, as his professional duty requires, because they are the larger group. However, he also knows that his romantic rival Ed is one of the passengers of the *Betty*. He hates Ed and would like to see him dead. This motive contributes to his desire to rescue the survivors of the *Amelia* rather than the *Betty*.

In *Coast Guard Rescue I* Brian’s motives that favor rescuing the *Amelia*’s survivors are mixed – he has one good desire (to do his professional duty by rescuing the largest number), and one bad desire (to have Ed die). But in this case I don’t think we conclude that Brian is blameworthy for his choice to save the *Amelia*’s survivors, even though he was led to act in part by his bad motive. His good motive renders him non-culpable. Perhaps this judgment is confirmed by noting that the facts of the story are consistent with its being the case that, had Brian believed Ed to be on the *Amelia*, his good desire would have outweighed his bad desire: he still would have chosen to rescue the *Amelia*’s survivors, including Ed, rather than the *Betty*’s survivors. We may, of course, evaluate Brian negatively for having the morally objectionable desire that his rival die, but this evaluation is not the same as an evaluation that he is overall blameworthy for his choice of how to act. Individuals frequently make choices from a mixture of motives, but the mere occurrence of a bad desire among good ones hardly makes these individuals blameworthy for the choices that issue from such mixed motives.

Thus inclusion of a bad desire is neither necessary nor sufficient for a configuration of motives to count as morally reprehensible. The reprehensibility of the configuration must emerge from the overall character of the set of desires. In some cases a configuration is reprehensible even though no component desire in itself counts as morally objectionable, and even though no good desire is lacking. Consider *Cheating*:

*Cheating*: Leila is worried that she may fail her organic chemistry test, since she has spent far too much time this semester worrying about a dysfunctional romantic relationship rather than studying. Because she wants to get into medical school, she cheats on the test, even though she cares about fairness and recognizes that it is unfair for her to cheat.

---

Neither of Leila’s desires (to promote self-interest and to do what is fair) is bad in itself. Nor does she lack a motive that she ought to have, since she does care about fairness.\textsuperscript{35} What is objectionable is the \textit{comparative strengths} of her motives: she chooses the wrong act because her motive of self-interest is stronger than her motive of doing what is fair. Thus it is the overall configuration of her motives (including their relative strengths), not any motive by itself, that makes her act blameworthy.\textsuperscript{36}

It might be said that Leila does have a bad motive: she is willing to do what she believes to be unfair in order to promote her self-interest. But I think it is a mistake to consider this as a motive as such. To say that Leila is willing to do what she believes to be unfair in order to promote her self-interest is just to say, using different terminology, that her desire to promote her self-interest is stronger than her desire to do what is fair. Alternatively it is a description of the decision that emerges from her configuration of motives. But it is not to attribute to Leila a distinctive desire different from the psychological states we have described already.

Our conclusion so far is that it is the moral character of the agent’s overall configuration of desires that renders her choice blameworthy, and that this moral character can emerge from the relative strengths of the desires that lead to her choice, rather than from the presence of a morally bad desire, or the absence of a morally good desire. We should also note that the mere presence of a morally bad desire in the agent’s psychology is irrelevant to the blameworthiness of the agent’s choice if that desire plays no role at all in generating the choice. Consider \textit{Coast Guard Rescue II}:

\textit{Coast Guard Rescue II}: Two boats, the \textit{Amelia} and the \textit{Betty}, have capsized in a storm. As he approaches the scene Coast Guard Captain Brian sees that he will have to choose to rescue one group of survivors or the other, and that members of the group he cannot rescue first are doomed to die of exposure, since he will not be able to get to them in time. He estimates that the passengers of the \textit{Amelia} are the larger group, so he heads his cutter in that direction, aiming to save them, as his professional duty requires, because they are the larger group. Brian’s mind is greatly occupied with his romantic rival Ed, whom he would like to see dead. Indeed, if Brian believed Ed to be among the \textit{Amelia}'s passengers, he would choose to rescue the \textit{Betty}'s passengers instead. However, since he doesn’t believe Ed to be a passenger on either boat, his desire for Ed to die plays no role in his choice to rescue the passengers of the \textit{Amelia}.

\textsuperscript{35} Someone might argue that she does lack a requisite moral desire, namely the desire to do what is all-things-considered right. But I don’t think we require non-blameworthy agents to have such a desire. Someone who chooses to help a frail person climb the stairs, simply out of concern for the person’s welfare, is praiseworthy, not blameworthy, even though she doesn’t think about morality at all in making her choice.

\textsuperscript{36} Note that there might be nothing unusual or deficient in the strength of her desire to do what is fair; it is just that her desire to promote her self-interest is stronger. Both of her desires might actually be weaker than such desires are in the average person. It is their \textit{relative} strength that is problematic. For an examination of the manner in which strength of desires may affect an agent’s moral status, see H. Smith (1991).
In this case Brian has an overt and morally objectionable desire, namely his desire that Ed die. However, since it contributes nothing to his decision to rescue the passengers of the *Amelia*, its existence is irrelevant to the question of whether or not Brian is blameworthy for his choice.\(^{37}\) Thus the desires that count, so far as the blameworthiness of a given choice goes, are only those that contribute to the occurrence of that choice. But we must include, among those that “contribute” to the choice, not only the desires that “favor” it, but also those that actively “disfavor” it. In *Coast Guard Rescue II* Brian’s hatred of Ed neither favors nor disfavors his choice to rescue the *Amelia*’s survivors — indeed it doesn’t motivationally connect with the choice at all — and so makes no contribution to the blameworthiness of his choice.\(^{38}\) By contrast, Leila’s desire to do what is fair “contributes” to her choice to cheat on the exam, even though this desire disfavors that choice. Because it plays a role, it may reduce her blameworthiness for the choice, even though the choice is overall one for which she is to blame. We must also remember that the lack of certain active desires (such as the instructor’s lack of concern for the impact of his theft on the graduate student) must be seen as “contributing” to his choice to steal the book. The relevant notion of “contributing to a choice” is clearly a complex one that merits deeper investigation.

I have claimed that the fact that an agent’s configuration of motivating desires includes a morally bad desire is not sufficient to show that he is blameworthy for the choice that issues from this configuration. Thus in *Coast Guard Rescue I*, Brian’s morally bad desire to see Ed dead, although it causally contributes to his choice to rescue the survivors of the *Amelia*, is not sufficient to show that his choice is blameworthy. However, there may be one special type of desire that is sufficient, by itself, to establish that an agent is blameworthy for his choice. Consider *Drug Gang*:

*Drug Gang*: Miguel and his young protégé Tomas, members of a Mexican drug gang, have kidnapped a local businessman. After securing a generous ransom payment, Miguel nonetheless decides to shoot and kill the businessman. He makes this choice partly from loyalty to Tomas, to protect him from possible future identification by the businessman, but also from the attractive thrill of doing something he believes to be all-things-considered morally wrong.

Miguel, like Brian in *Coast Guard Rescue I*, chooses on the basis of morally mixed motives: his desire to protect his protégé Tomas is arguably a good motive, while his desire to do what he believes to be all-things-considered morally wrong is clearly a bad motive. But, whereas we don’t think that Brian is blameworthy for his choice, we do think Miguel is blameworthy for his choice. And the reason he is blameworthy is that he is motivated by a desire to do what he believes to be all-things-considered wrong.

---

\(^{37}\) This is true even if Brian’s choice is blameworthy. If Brian believes that duty requires him to rescue the more numerous passengers from the *Amelia*, but chooses to rescue the passengers from the *Betty* instead because he believes one of them is a billionaire who will reward him handsomely, his choice is blameworthy, but not because he also has an overt desire that Ed die.

\(^{38}\) There may be desires that motivationally connect with the agent’s decision, but which weigh evenly on each side of the choice, so on balance they neither favor nor disfavor it. Nonetheless, because they motivationally connect with it, they are relevant to the agent’s overall blameworthiness.
Clearly the desire to do what one believes to be all-things-considered morally wrong is morally more “powerful” in establishing blameworthiness than other kinds of merely bad desires.\textsuperscript{39} It appears that this special desire is often sufficient by itself to establish blameworthiness for a choice, even if other kinds of morally bad desires are not sufficient.\textsuperscript{40}

Consideration of Drug Gang suggests something else. Some agents, such as Miguel, are motivated by desires that are explicitly moral in content. Other agents, however, don’t think about the moral character of their prospective actions when they make choices. Instead they are attracted or repelled solely by the underlying morally relevant features of their prospective actions – the very features that make them right or wrong. Consider Intern:

**Intern:** Shelly feels intense rivalry with her co-worker Greg. Greg supervises a new intern, Gina. Shelly discovers that Gina has made a fairly serious error in preparing materials for an important staff meeting. Shelly could privately point out the error to Gina and let her correct the error before the meeting, or she could quietly alert Greg so that he could point out the problem to Gina. Alternatively, Shelly could wait until the meeting and then call the error to everyone’s attention. This would publicly embarrass Gina, which Shelly feels bad about. However, it would also humiliate Greg by casting him in a bad light as a supervisor, which Shelly finds very attractive. Shelly decides to wait until the big meeting to point out the error.

Shelly has no thought about what is right or wrong in making her choice. She is simply attracted by the idea of humiliating Greg, while also being averse to the idea of embarrassing Gina. Most of us see her desire to humiliate Greg as a morally bad desire (even though it has no explicit moral content), and see her aversion to embarrassing Gina

---

\textsuperscript{39} Notice that I am interpreting *Coast Guard Rescue I* as a case in which Brian has no beliefs about, or motivation regarding, the all-things-considered moral status of his options. He is motivated solely by considerations of professional duty and hatred for his rival. However, we can easily imagine a variant of *Coast Guard Rescue I* in which Brian believes, and is motivated by the belief, that his rescuing the survivors of the *Amelia* is the all-things-considered morally right act. In this variant he would also not be blameworthy, since part of his motivation would focus on the all-things-considered status of his proposed action.

In this variant Brian desires to do something (let Ed die) that he recognizes as prima facie wrong, and this leads us to see him as morally flawed. But since he has no desire to do something that he recognizes as all-things-considered wrong, we don’t conclude that his choice is blameworthy.

\textsuperscript{40} We should also make room for a person being blameworthy for an act if he believes it to be all-things-considered subjectively wrong. I shall leave this complication aside.

As Kenneth Simons points out, this statement may be too strong, since there may be cases (a teenager who gets a thrill from violating social conventions and moral norms) in which we wouldn’t be clear that the behavior was blameworthy. Certainly the degree of blameworthiness can vary with the importance of the norm intentionally violated, and there may be cases in which we don’t view the behavior as blameworthy at all, if the norm is a sufficiently modest one (Simons’ examples: showing politeness, not talking back to one’s parents, etc.)
as a good desire (even though it has no explicit moral content). The first of these is a desire that it is morally better not to have, while the second is a desire that it is morally good to have. Unfortunately in Shelly’s case the former desire is stronger than the second, so we judge her to be blameworthy for her choice. This suggests that agents can be blameworthy for their choices even though their desires have no explicit moral content. It further suggests something more complicated: that an agent is blameworthy for her choice if her configuration of non-moral desires has a certain character, namely that the desires are responses to supposed features of the prospective act which, taken together, would qualify the act as all-things-considered wrong. We might formulate this as follows:

III. S is blameworthy for her choice to perform act A even if S does not believe about A that it has any all-things-considered deontic status, just in case:

a. S believes that A would have features F, G, and H,

b. S’s choice to perform A arises solely from her desires or aversions (or lack thereof) with respect to A’s having F, G, and H, and

c. If A occurred and had F, G, and H (and no other morally relevant features), then A would be all-things-considered wrong.

There are several ways in which an agent can go morally astray in responding to features of an action that would make it all-things-considered wrong. Suppose that F and G are actually right-making features, while H is actually a wrong-making feature. The agent might respond appropriately to the *valence* of these features: she is attracted by F and G, and repelled by H. However, her response is not appropriate to the *strength* of one or more valence. For example, she might be much less averse to H than it warrants, so that in her psychology, her desires for F and G inappropriately outweigh her aversion to H. Leila, in *Cheating*, is an example of this. She is appropriately averse to doing what is unfair, and legitimately wants to do what is in her self-interest, but unfortunately she is not sufficiently averse to doing what is unfair. Alternatively, the agent might respond inappropriately to the valence of these features: perhaps she is attracted rather than averse to H (or lacks any conative attitude towards H), and so makes a choice that would be all-things-considered wrong. Shelly, in *Intern*, is an example of this; she is attracted to the idea of humiliating Greg, when she should be averse to it.

What these considerations show, I believe, is that an agent’s blameworthiness for an act must reflect her overall assessment of the act, where her “overall assessment of the act” can consist either of her evaluative attitudes towards the act’s all-things-considered moral status, or her evaluative attitudes towards supposed features of the act that would

---

41 Of course a strict Kantian would deny that any “motivation” aside from respect for duty has moral significance, but few people are strict Kantians about this issue.

42 I am interpreting “moral” desires as ones with explicit moral content, such as the desire to do *what one is obligated to do*. “Non-moral” desires are all other desires, many of which of course are morally important, including the kinds of desires described in the text.
determine its all-things-considered moral status. We can combine these ideas in the following account.\(^{43}\)

IV. S is blameworthy for her choice to perform act A just in case either:

A. S believes that A would be all-things-considered wrong and chooses A; or

B. Although S lacks any belief about A’s overall deontic status,

1. S believes A would have features F, G, and H,
2. S’s choice to perform A arises solely from her desires or aversions (or lack thereof) with respect to A’s having F, G, and H, and
3. If A occurred and had F, G, and H (and no other morally relevant features), then A would be all-things-considered wrong.\(^{44}\)

On this account what makes a choice blameworthy echoes, in an important way, what makes an act all-things-considered wrong. Many acts have wrong-making features, but having a wrong-making feature does not make an act wrong all-things-considered, since the act’s right-making features may outweigh its wrong-making features. It is not wrong all-things-considered for a veterinarian to break your dog’s leg, even though this aspect of his act is a wrong-making feature of it, if his act also has the weightier right-making feature that it will enable him to re-set the leg which has healed badly from an earlier injury. It is the overall configuration of the act’s morally relevant features that

\(^{43}\) A similar account could now be stated for an agent’s being blameworthy for performing an act. Note that these accounts do not overtly cover cases in which the agent knowingly risks causing a harm. Such cases can be incorporated by understanding “act A” so that it can be the act of “risking a harm,” and understanding “features F, G, and H” so that they can be probabilistic features, such as “is likely to lead to an injury.”

This account would have to be complicated in order to correctly handle cases in which S believes that A would be all-things-considered wrong, but also that A has the highest expected value of any of S’s alternatives and is therefore subjectively right. See (Zimmerman 2008, pp. 17-18, and H. Smith 2011).

\(^{44}\) Note that this account doesn’t satisfactorily handle “Huck Finn” type cases, in which the agent both believes his act to be all-things-considered wrong, and also believes it to have features F, G, and H, in virtue of which he wants to perform the act – but it is also the case that if the act had features F, G, and H, then A would be all-things-considered right. Huck Finn famously believed it would be all-things-considered wrong for him not to turn in his friend Jim, who is a slave, and also wanted not to turn in Jim because of Jim’s humanity. People’s reactions to such cases understandably vary; some respond that Huck is blameworthy for choosing to do what he believes to be all-things-considered wrong; others respond that Huck is praiseworthy (and certainly not blameworthy) for choosing to do what, for the non-normative reasons he chooses it, is all-things-considered right. Here I will not take a position on such cases, but only note that one’s final account of when an agent’s choice is blameworthy must take such cases into account, as Account IV does not attempt to do. For one discussion of this case, see Arpaly (2003), pp. 75-8 and passim.
make it all-things-considered right or wrong. Moreover, it is not blameworthy for the vet to break the dog’s leg in this case, even if he believes that breaking a dog’s leg is prima facie wrong, because the vet’s blameworthiness must be connected to his view about the overall moral status of the act, not to his view about the moral status of just one of its features. It should not be surprising, then, that in order to establish her culpability for a choice, the agent’s psychology must in some way reflect a response to the overall moral character of the act. One way to reflect this character is for the agent’s psychology to include a desire to perform the act even though she views it as all-things-considered wrong. But another way to reflect this character is for the agent’s psychology in making the choice to exhibit an overall configuration of motives that is morally flawed taken as a whole – a notion that can be spelled out as in Clause B of Account IV. Simply having one bad desire, as does Brian in Coast Guard Rescue I, is not sufficient to show that he is blameworthy for the choice that is partly motivated by this desire – just as an act’s having one wrong-making feature is not sufficient to show that it is wrong all-things-considered.

D. A further complication in blameworthiness

I have argued that what makes an agent blameworthy for a choice is the character of her overall evaluative attitude towards what she believes to be the chosen act’s features. However, we need to consider another aspect of this situation. Consider Bad Hair Day:

**Bad Hair Day:** Clara dislikes her classmate Bonnie, and has always thought Bonnie’s hair style to be really unattractive. She has often been tempted to wound Bonnie by saying something nasty about her hair. However, because Clara doesn’t want to get a reputation for being mean, and wants to impress her new boyfriend with what a nice person she is, she has always resisted the urge to make a nasty comment about Bonnie’s hair. During an experiment Clara’s psychology teacher hypnotizes Clara. Part of the experiment (unbeknownst to Clara) involves inhibiting, for twenty-four hours, certain of Clara’s desires: her concern to maintain a good reputation, and her desire to impress her boyfriend with her good character. That evening, while cruising social networking sites, Clara comes across a picture of Bonnie on her Facebook page. Clara’s desire to wound Bonnie is aroused. Since her other pertinent motives are “frozen” as a result of the hypnosis, Clara decides to post a cutting attack on Bonnie’s appearance. She knows the attack will be read by all their mutual acquaintances, and deeply humiliate Bonnie.

Is Clara blameworthy for deciding to launch the attack on Bonnie? According to Account IV, her choice is blameworthy, since her choice issues from a single reprehensible desire, namely the desire to wound Bonnie. An act whose only morally relevant feature is that it will wound someone is an act that is all-things-considered wrong. Since Clara is solely motivated by this feature, our account deems her choice to be blameworthy. Of course Clara’s choice is affected by her being hypnotized. But in this case, unlike in many philosophical hypnotism cases, the desire to wound Bonnie is genuinely one of Clara’s desires; it is not a “foreign” desire that the hypnotist implants in
However, contrary to the implication of Account IV, I think deeper reflection leads us to assess Clara as not blameworthy. And this seems to be precisely because we understand that Clara has other motives that would normally have contributed to her making a decision about posting the screed, and that these desires did not play their usual role once she was hypnotized. In other words, although her choice did arise from her own desire, and does reflect her entire evaluational response at that moment to her options, the choice did not arise from anything like a reasonably full configuration of the motives that she actually has and that would normally bear on such a decision. Her choice reflects part of Clara’s psychology, but not enough of her psychology to warrant a judgment of blameworthiness. To blame someone is to negatively evaluate that person’s whole motivational structure insofar as it bears on the choice, and Clara’s choice does not reflect anything like her full personality.

Of course hypnotism cases are complicated. Even though Clara acts from one of her own desires, nonetheless she is prevented from responding normally to her situation by the intervention of another agent, namely the hypnotist. We may see this intervention as interfering with her freedom, and therefore as diminishing her responsibility in the case. Perhaps this aspect of the case, rather than the fact that Clara’s response doesn’t affect her full motivational structure, is what governs our denial that she is blameworthy for her choice.

However, hypnotism is not the only process that prevents significant components of an agent’s motivational structure from playing their normal role. Contemporary research in cognitive science has shown that normal individuals are often caused to behave in certain ways by environmental stimuli while remaining completely unaware of the role these stimuli play in causing their behavior. Here is a report of such an experiment:

New York, 1996: University students take part in an experiment on the effects of behavior-concept priming. As part of an ostensible language test, participants are presented with many words. For some participants, words synonymous with rudeness are included in this test; for others, words synonymous with politeness are included instead. After finishing this language test, all participants are sent down the hall, where they encounter a staged situation in which it is possible to act either rudely or politely. Although participants show no awareness of the possible influence of the language test, their subsequent [rude or polite] behavior in the staged situation is a function of the type of words presented in the test (Bargh 2005, p. 37).

Well-known experiments have further shown that when older adults are subliminally presented with terms associated with being elderly, their memory performance degrades if they have been presented with negative terms such as “senile” and “dementia,” while their performance improves if they have been presented with positive words such as “wise” and “experienced” (Choi et al. 2005, p. 320). Similarly,

---

45 See, for example, Wolf’s discussion of hypnosis as undermining autonomy when the hypnotist “implants” a desire (Wolf 1990, p. 9). Scanlon (1998) echoes this (pp. 277-8).
when study participants are subliminally exposed to happy, neutral or angry facial expressions, their faces subtly mimic the subliminally presented faces, of which they are not consciously aware (Choi et al. 2005, p. 314). Further studies show that people unconsciously mimic the posture and physical gestures of interlocutors (Chartrand and Bargh 1999, pp. 893-910), and that subtly activating a stereotype of elderly persons causes college students to walk more slowly (Dijksterhuis and Bargh 2001). People who are exposed to a foul smell report more negative attitudes towards gay men on a survey than people who answered the survey in the absence of the stench (Liberman and Pizarro 2010). In some of these cases, the individual is not only unconscious of the causes of her behavior, but also unconscious of the behavior itself.

In discussing these findings John Bargh posits that many if not most of the motor systems that guide actions are opaque to conscious access, and that much of our behavior is guided by automatic processes that are not governed by conscious control and guidance. Various characterizations of the “automaticity” at issue have been offered, but one description characterizes such psychological processes as “automatic” in the sense that they are inevitably invoked by environmental stimuli; that once triggered, execution proceeds rapidly, effortlessly, and incorrigibly to completion; and the process leaves no traces accessible to conscious recollection (Glaser and Kihlstrom 2005, p. 171). Having the capacity to develop and utilize such automatic processes, Bargh speculates, is highly beneficial, since it enables us to devote our highly limited conscious capacities to decisions and matters that require conscious attention, leaving more routine matters to the unconscious processes (Bargh 2005, pp. 43, 53).

The question for us is whether or not an agent who acts, or chooses to act, as a result of an automatic process of the sort just described is blameworthy when her proposed action is wrong. Of course, most actions elicited by psychologists in the pertinent experiments are not wrongful actions. Nonetheless some can be seen as being at least minimally wrong. For example, the students who respond rudely in the staged situation may well be acting wrongly. In another experiment, subliminally activating the African-American stereotype (which, according to the experimenter, includes hostility) through subliminal exposure to faces of young Black men causes young white research subjects to react with greater hostility to an experimenter’s request (Dijksterhuis and Bargh 2001). These hostile responses, too, could well be wrongful. In some of the cases the acts (or decisions) are clearly wrong: research subjects who were primed with “hostility” gave more intense electric shocks in Milgram-type experimental situations (Dijksterhuis and Bargh 2001, p. 16). Thus these experiments appear to provide real-life, sans-hypnosis, examples of what we imagined in Clara’s hypnosis case: individuals

---

46 Cited in Bargh 2005.
47 Glaser and Kihlstrom dispute the accuracy of some of these descriptors, in particular citing research to show that the unconscious can compensate for threats to the achievement of its goals.
48 For further research on dual process theory, see Evans and Frankish (2009). While Hassin et al. (2005) report research among social psychologists, Evans and Frankish report research across a broader range of fields. For a penetrating discussion of the extent to which the kinds of experiments cited genuinely establish unconscious operations of the mind, see Uhlmann et al. (2008).
whose behavior is triggered by an environmental stimulus in a way that doesn’t engage their full appraisal of the situation or the appropriateness of their act.

Is an agent who chooses to perform a wrongful act as the result of such an automatic process blameworthy for his choice? One might be tempted to say “no” on grounds that the desire that triggers his action – for example, a desire to respond with hostility to an experimenter – is not genuinely one of the agent’s own desires, but instead is “implanted” in the agent by the experimenter. However, this response seems inadequate. For one thing, there need be no experimenter or other calculating human agent behind the environmental situation that triggers the desire. Merely glimpsing a young Black man on the street might be enough to trigger a hostile response in the next encounter, and merely seeing an elderly person in a hotel, or reading a newspaper article on retirement, might trigger the desire to walk more slowly. Indeed, the kinds of environmental triggers studied in these experiments are the kinds we encounter all the time in daily life, and it is highly plausible that we often react in a similar manner.49 Second, what we regard as our “normal” desires are also often triggered by environmental stimuli. My desire to eat a donut is triggered by the smell wafting from a Dunkin Donuts shop; my desire to walk on the other side of the street is triggered by seeing a fierce dog on the sidewalk in front of me; my desire for a cold beer is triggered by seeing a beer advertisement. Of course one might say in the latter cases that I must already have underlying generic desires – hunger, fear of dogs, thirst – which are genuinely mine, and which combine with the environmental stimuli to trigger my specific desires for a donut, to drink a beer, and so forth. However, this may be true of the automatic cases also. Perhaps I only acquire the desire to walk slowly on seeing an elderly person or to cross my legs when someone across from me crosses her legs because I have an underlying desire to conform myself to others around me. But in either case the specific form of the desire, whether it’s a desire to eat a donut or to cross my legs, is triggered by the environment. If these mechanisms are as pervasive as psychologists suspect they are, many of the desires that activate our behavior arise through just such stimuli. There seems no ground here to reject any of these desires as not being “really” the desire of the agent.50

Nonetheless there are two kinds of cases here. In many of the “automatic” cases (such as crossing one’s legs when one’s interlocutor crosses hers) I believe that the person is not responsible for what she does in response to the environmental trigger, in

---

49 Witmer (2001) provides a poignant example of this. His daughter Michelle was the first female National Guardsman killed while serving in Iraq. Witmer writes that on hearing the news of her death, his first instinct was to bring his other two daughters home from their military assignments. He notes wryly that his first instinct was to protect his daughters, not to think about the fact that they were adults who had to decide for themselves, or to think about the patriotic need to support the country (John Witmer 2010).

50 Discussions of whether an act is one for which the agent is responsible often raise the question of whether the act is truly “free” because the desire that gives rise to it comes from “outside” the agent, as in cases in which a hypnotist or brain surgeon implants a foreign desire in an agent, who then acts on it. There are serious issues about whether or not such agents are responsible for their acts, but the question is irrelevant to our concern, since the attitudes and desires in most of my cases are not foreign to the agent in this sense.
contrast to the person’s being responsible for what she does in a non-automatic case (such as buying a donut) in response to the environmental trigger. The reason for this is that the “automatic” act in question is triggered straightway by a single desire (or a highly limited set of desires), and the agent’s performing it does not reflect any of the agent’s wider psychology – just as Clara’s launching the Facebook attack while hypnotized does not reflect her wider psychology. The desire to cross my legs to mimic my interviewer is a desire of whose origin I am unaware. In some cases I may not even be aware of the desire. The behavior does not reflect my all-things-considered desire to mimic my interviewer, since it has not had to compete with other, possibly countervailing, desires, or to survive conscious scrutiny. Nor does it appear to emerge from some unconscious competition or balancing act among my non-conscious desires. The behavior simply emerges directly from the desire, and the desire itself is triggered directly by my perception of my interviewer’s posture, without other features of my situation playing any significant role.

By contrast, in the donut-buying case, while I may or may not be aware of what caused my sudden desire for a donut, I am aware of the desire, and aware that I have the option to buy one or not, and these facts result in a weighing of my desire for the donut against other desires: my desire to maintain my weight, or my desire to avoid sticky fingers when I’m on my way to a job interview, or my desire to acquire an even better donut from another donut shop. Typically, if I do buy the donut, my decision to purchase it reflects my broader consideration of my reasons for and against, even though the desire itself is directly triggered by an environmental stimulus.

It is unclear to me how crucial the role of consciousness is in drawing this distinction between automatic processes and the more reflective processes that involve weighing a desire against other desires. In the cases in which you are aware of a desire, and aware that satisfying that desire would involve acting in a certain way, then you typically do weigh satisfying that desire against other considerations. If you are not aware of a desire (e.g., the desire to walk more slowly because you’ve heard the word “elderly”), then you are less likely to weigh the importance of the desire against other considerations you deem relevant. But it would take further psychological research to show whether or not the comparison or weighing process that normally takes place in consciousness could also take place at the unconscious level. For our purposes the important point is that in the kinds of automatic cases described so far, for the most part it seems as though the process leading from the triggered desire to the action is a process that is more or less isolated from the agent’s other motivations, and so does not adequately reflect the agent’s full psychology, or full evaluation of the act to be performed.\footnote{Of course the situation could have been different. If a friend had warned you to, at all costs, avoid crossing your legs during the interview, because the interviewer is offended by women crossing their legs, then you would have been alert for any inclination to cross your legs, and would have suppressed it. However, without any such warning, your guard is not up.}

If this is correct, it suggests that choices reached as the result of such automatic processes are, like Clara’s choice, ones for which we should not blame the agent.
Although the choice fully reflects the agent’s *active* evaluative attitudes towards her prospective act, and so qualifies as blameworthy according to Account IV, it fails to sufficiently reflect anything close to her full set of evaluative attitudes towards the prospective act. We may evaluate the desire that gave rise to the choice as morally objectionable, and we may evaluate the agent as morally flawed because she has that desire, but we should not conclude that the agent was overall blameworthy for making the choice triggered by that desire. Thus Account IV of blameworthiness needs to be revised to accommodate this point.

What we are discovering is that blameworthiness for a choice (when the agent is not motivated by a desire to do something she regards as all-things-considered wrong) depends on the full set of desires and aversions that give rise to the choice – but also that blameworthiness is only appropriate if that set of desires and aversions itself constitutes a sufficiently complete range of the pertinent motivations of the agent. We need to be cautious here. It is probably very rare for a decision to perform an act to reflect the agent’s *full* psychology. Although we have been looking at other kinds of psychological processes, we should not forget that ignorance and mistake play an important role here. Agents have many desires, indeed many desires that are pertinent to the decisions they confront, in the sense that the decision may have upshots that either satisfy or frustrate these desires. However, if the agent is not aware that his decision is relevant to one of his desires, the desire does not influence the decision. This is one important way in which ignorance blocks culpability. Still, we don’t want to say that an agent is only blameworthy if his *entire* set of pertinent desires plays a role in his decision. It is not easy to see where to draw the line between an agent who is blameworthy because enough of his motivational structure was engaged by the decision, and an agent who is not blameworthy because *not* enough of his motivational structure was engaged by the decision. Probably there is a continuum here: we should speak instead of *degrees* of blameworthiness, where an agent is more blameworthy if a substantial portion of his psychology was engaged in a decision, and less blameworthy if a less significant portion

---

52 Note that the number, as such, of the agent’s desires and aversions, is not the relevant factor. If an agent has only one relevant (bad) desire that could be called into action by a prospective act, and indeed that desire does, by itself, lead to the agent’s decision to perform the act, then we can judge the agent to be blameworthy for her decision, even though it arose from a single bad desire. What is relevant here is not the number of desires that give rise to the decision, but the extent to which they represent the agent’s full set of attitudes towards the act in question.

53 Theorists sometimes say that an agent’s ignorance blocks the otherwise natural inference from the fact that the agent performed a wrong act to the conclusion that the agent had a reprehensible desire to perform that act. What I am pointing out is that ignorance operates to block culpability in another way as well, by blocking the inference that a reprehensible desire actually possessed by the agent played any role in contributing to the agent’s decision.

54 Many theorists of blameworthiness have claimed that a person’s blameworthiness depends on whether or not his act or his choice arises from his character. This seems incorrect, since a person’s character is a stable psychological feature, and some acts seem to be blameworthy even though they arise from fleeting motivational states that are too short-lived to count as part of the person’s character. However, the impulse to look for states of character in assessing blameworthiness may arise from an unconscious recognition that blameworthiness requires that the action or choice issue from something close to the agent’s full psychology at the time of the choice.
of his psychology was engaged. However, it seems reasonable to say that a choice triggered by a *single* desire, when the agent has other pertinent desires that are not activated and so play no role, involves too little of the agent’s motivational structure to render him blameworthy for the decision. A judgment about blameworthiness is a judgment about his overall pertinent motivational structure at the time of choice, and the activity of a single bad desire may not tell us enough about his overall motivational structure to license such a judgment.

Although ignorance and mistake often play an important role in preventing an agent’s full array of pertinent desires from affecting his decision, the psychological experiments I have described show us that other psychological processes can be just as important in preventing pertinent desires from affecting an agent’s decision. These experiments are reasonably interpreted as cases in which the environmental stimulus directly causes a desire that in turn directly triggers a decision, without the activation of further desires or aversions that might be relevant to the act. Even without the insights gained from formal psychology experiments we know that decisions triggered by single desires, without any input from other motivations, occur, and we hesitate to hold their agents blameworthy, or fully blameworthy. One species of such decisions are ones we might call “hair trigger” decisions. They include such decisions as a soldier’s decision to fire his gun at an unidentified noise when he is in a very dangerous environment and has been trained to shoot first and ask questions later; or someone’s decision to throw up his hands to protect his face from an object unexpectedly thrown at him from close quarters; or someone’s automatically slapping at a mosquito; or someone’s saying “ouch” at the sudden twinge of a bad knee. Unlike true reflexive actions, these actions are governed by central brain processes, and are not in principle outside the agent’s control: in hair-trigger acts the agent could have modified his reaction if he had had different training or had enough forewarning. But when he hasn’t had such training or forewarning (and isn’t culpable for this fact) we should not hold him morally responsible for his decision, which in some ways is more like a reflex: he responds before thought, or broader consideration, has the opportunity to kick in. Indeed, we find it unnatural to describe such cases as ones involving a “decision” at all, because we think of decisions as involving greater reflection, and consideration of various aspects of the choice, than occurs in such cases.

Such hair trigger cases have a very different moral tone from ones in which many of an agent’s relevant motives come to bear on his decision, but a single morally flawed motive overwhelms competing morally good motives. These “overwhelming” type cases are

---

55 My colleague Doug Husak is persuading me that the notion of degrees of culpability is what is appropriate here.

56 Nor does the agent have the thought, on deciding, that this act is all-things-considered right or wrong. It would be possible to have such a belief without having consulted more than one reason for acting, but the agents in “hair trigger” cases typically do not decide on the basis of this thought (which might be enough to establish blameworthiness or lack thereof).

57 Someone might say that it is part of the agent’s “full moral personality” that his other pertinent motivations are *not* activated in such cases, so that we can hold him blameworthy. However, this would require us to hold that the failure of the triggering stimulus to recruit his other motivations shows his configuration of desires and aversions is reprehensible. This is implausible, since it is simply part of the way human beings are “wired up” that we respond in these ways to these kinds of stimuli. And, as John Bargh argues, such wiring is evolutionarily extremely important for survival (Bargh 2005, pp. 43, 53), so we do not want to claim that it is morally reprehensible.
ones in which we properly hold the agent blameworthy, unlike the cases in which the potentially competing motives are never even activated.

E. Blame for non-voluntary responses

We have taken a long detour to become more clear on what conditions render a person blameworthy for a paradigmatic decision to act. Our conclusion can be encapsulated in the following revised account of blameworthiness for a decision:

V. S is blameworthy for her choice to perform act A just in case either:

A. S believes that A would be all-things-considered wrong and chooses A; or

B. Although S lacks any belief about A’s overall deontic status,

1. S believes A would have features F, G, and H,
2. S’s choice to perform A arises solely from her desires or aversions (or lack thereof) with respect to A’s having F, G, and H,
3. If A occurred and had F, G, and H (and no other morally relevant features), then A would be all-things-considered wrong, and
4. S’s desires and aversions (or lack thereof) with respect to F, G, and H represent a sufficiently complete set of her desires and aversions that are relevant to act A.58

58 More work needs to be done on the concept of “relevance” in this account. In the normal case, a desire is relevant if the prospective act has feature J, and it is true that if S believed the act to have J, and if S’s motivational system were fully active, then S would desire to do the act because it has J (or would be averse to the act because it has J), and that desire would be active in S’s decision whether or not to perform A. For example, suppose in Coast Guard Rescue II that Captain Brian’s rival Ed is on the Amelia, but Brian does not believe Ed to be on the Amelia or that his rescuing the Amelia survivors will involve saving Ed’s life (J). Then Brian’s desire to see Ed dead is “relevant” to his choice to rescue the Amelia’s passengers, even though his ignorance prevents this desire from playing an active role in his choice.

However, S will sometimes have inaccurate beliefs about A, beliefs that would normally trigger a desire with respect to A, but which (because it is an automatic case) are not engaged by the choice to do A. Thus S might falsely believe A to have feature K, and have a standing desire not to do acts that have K – but because S’s decision to do A is an automatic decision, his belief and desire with respect to K have no impact on his decision. For example, suppose the soldier in the dangerous environment falsely believes that shooting under these circumstances will disobey his commander’s orders (K). (His belief is false because the commander, bent on revenging a comrade’s death, has not issued such orders for this situation.) Normally the soldier would desire to obey orders, but because of the perceived danger his desire to obey orders is never activated. Since the soldier’s desire to obey orders would normally be an active part of his overall evaluative attitudes with respect to shooting (as he understands this act), it seems that it should count as a relevant desire as well, even though shooting would not in fact disobey his commander’s orders.
Doubtless this account needs a good deal of refinement, including more detail about the crucial concept of “sufficiently complete set of desires and aversions.” However, it is adequate for our immediate purpose, which is to draw lessons from the conditions required for a blameworthy decision for the parallel case of a non-voluntary response of the kind an Attributionist wants to hold blameworthy – a response such as a spontaneous emotional or cognitive reaction that reflects the individual’s evaluative judgment, or the non-voluntary outward signs of these inward states such as blushing or wincing.

If a person is to be held blameworthy for a non-voluntary response, it appears that she must meet conditions analogous to those she must meet in order for one of her decisions to be blameworthy. By making suitable adjustments to Account V, we might propose the following parallel account of blameworthiness for non-voluntary responses:

VI. S is blameworthy for her non-voluntary response R to situation X just in case:

A. S believes X has F, G, and H,
B. S’s response R arises solely from her desires or aversions (or lack thereof) with respect to X’s having F, G, and H,
C. If X had F, G, and H (and no other morally relevant features), then response R to X would be all-things-considered bad, and
D. S’s desires and aversions (or lack thereof) with respect to F, G, and H represent a sufficiently complete set of her desires and aversions which are relevant to situation X.

Thus consider the politically liberal person who reacts, to her surprise, with a pang of dismay on learning that her new dentist is African-American. This dismay reveals a genuine (hitherto hidden) racist attitude on her part, and we (and she) may appropriately judge her as morally flawed because of her bigoted attitude. She would be a better person if she lacked this attitude. But should we judge her to be blameworthy for this response? The response reflects a reprehensible evaluative attitude, and the response itself is indeed morally inappropriate to the features of the situation. Nonetheless, according to Account VI, we should not judge her as blameworthy, because her spontaneous response does not reflect enough of her psychology. Her full configuration of attitudes relevant to her dentist may include admirable ones, such as positive attitudes towards equality among the races, views that it is wrong to denigrate persons because of their race, hopes that people will be judged by their qualifications rather than their race or gender, and so forth. The fact that this mix includes a “gut-level” racist response to the

---

59 Note that response R must be appraised as “bad” rather than “wrong,” since R is not a voluntary act.
60 Some racists would have a reaction of dismay because of ignorance or mistake – for example, they might believe that African-Americans do not have the intellectual capability to become good dentists. This ignorance or mistake may itself be culpable, but such a case would involve a “tracing” instance of culpable ignorance. I am imagining a racist who now has no such beliefs (although she may have held them in the past).
prospect of being treated by an African-American dentist is unfortunate, but it is not anything like the whole story about her evaluative attitudes towards African-Americans.\textsuperscript{61} According to Account VI, then, it does not provide grounds for us to hold her to be blameworthy for the response.

Although there may exist non-voluntary responses that reflect the agent’s full moral personality, it appears to me that in the majority of cases such a response issues from a highly limited set of the agent’s concerns, and in many cases, a single one of her concerns. Most of the research I have cited has focused on agents’ acting in “automatic” ways, rather than on agents’ experiencing non-voluntary responses such as blushes and feelings of dismay. Nonetheless some of the experiments involve agents’ experiencing mental states (good or poor memory performance; evaluations of gay men), and others involve typical non-voluntary responses (facial expressions). Although this is a matter for more empirical research, reflection on these experiments, as well as inspection of one’s own case, makes it seem highly plausible that many, if not most, of the non-voluntary responses of this sort are typically triggered by a single evaluative attitude, not by the full configuration of the agent’s overall evaluative attitudes. One’s obsessively dwelling on the last donut is caused by greed. One’s other attitudes at the time – the desire not to appear piggish, the desire not to gain weight, the desire not to take more than one’s fair share, etc. – play no role in causing (or in inhibiting) the non-voluntary response of thinking about the donut, although these other attitudes certainly play a role in preventing one from actually reaching for the donut. The non-voluntary response is directly caused by greed alone. Similarly your blushing is caused by your embarrassment. Your other attitudes – your desire to maintain dignity, your desire not to let someone know he has embarrassed you – play no role in causing (or inhibiting) the

\textsuperscript{61} Her reaction may arise from what Tamar Szabo Gendler has labeled “aliefs”—epistemic states that one retains in the face of overwhelming evidence to the contrary. See, for example, Gendler (2008), in which she describes experiments in which subjects refuse to drink soup from a brand-new bedpan, are loathe to put their mouths to a piece of vomit-shaped rubber, and are averse to tasting fudge made in the shape of dog feces (p. 636). These subjects appear to have “belief-like” states (e.g., the “belief” that a bedpan is contaminated) which give rise to their emotional reactions, despite the fact that they have contrary genuine beliefs (the knowledge that the bedpan is not contaminated).

What should we say about the response of what we might call a “confirmed” racist who evinces the same pang of dismay on learning that his new dentist is African-American? A “confirmed” racist is someone who has thought about racism and, for bad reasons, accepted racist attitudes as appropriate. For this reason he has made no attempt to eliminate his own biases. The “confirmed” racist’s response to the new dentist, like that of the agent in the text, is also non-voluntary. However, his response traces back to a history in which he has culpably failed to extirpate his racist attitudes. Thus his case should be considered a “tracing” case, in which he is blameworthy for his current non-voluntary response because it can be traced back to his earlier failures. The agent in the text also thought about race in the past, but her consideration was not flawed, and so cannot infect her current response with culpability.

Of course we can imagine a case in which someone’s non-voluntary response to a stimulus is caused by an objectionable evaluative attitude, and the objectionable evaluative attitude is part of an objectionable set of evaluative attitudes on her part – but it is the only attitude activated by the stimulus in question. Hence her non-voluntary response does not tell us the full story about her moral psychology, although the full story is not a good one. My view implies that in this case, as in the case of the liberal racist in the text, the person is not to blame for her non-voluntary response, since it does not sufficiently reflect her full psychology.
non-voluntary response of blushing. Thus there is typically a direct and immediate connection between a single evaluative attitude and your non-voluntary response, but the non-voluntary response in no way is governed by your full set of evaluative attitudes. In this important respect your non-voluntary response can show us (and you) that you have an objectionable evaluative attitude. It tells us something about you, and perhaps the news isn’t favorable. But it doesn’t show us anything about your all-things-considered evaluation of the situation that triggered your non-voluntary response. Just as we don’t blame the person who harbors a bad motive but acts rightly from a more complex consideration of her action, so we shouldn’t blame a person who has a contemptible evaluative attitude, but merely displays that attitude through some non-voluntary response that indicates nothing about what her overall evaluative attitude is towards the situation that gives rise to the response.

It appears, then, that following the logic of our discussion of blameworthiness for decisions, we must conclude that there are many occasions on which people cannot legitimately be blamed, although they can be criticized as exhibiting a moral flaw, for their non-voluntary responses. Deciding, in any particular instance, whether a non-voluntary response has this character, or instead reflects more fully the agent’s views about the situation, is not necessarily an easy practical task. But it is an important one if we are to place blame accurately where it lies.

This discussion draws attention to what may be a main difference between a judgment that a single attitude is morally objectionable and a judgment that a person is blameworthy for having that attitude. Philosophers have wanted to draw the distinction between negatively evaluating a single attitude (or saying a person is “bad” for having that attitude) and judging someone as blameworthy for that attitude, but their accounts have not been wholly illuminating or persuasive (Watson 2004, Levy 2005, and A. Smith 2008). The difference I am pointing to here may well be the key.

F. Blameworthiness for failures to notice

Let us now bring this discussion back to our original question about culpability for non-tracing cases of failures to notice. In the cases with which we are concerned, the agent’s failure to notice some important feature of his circumstances leads him to act wrongly. It is claimed, and we have accepted this claim at least for purposes of argument, that in some of these cases the agent’s failure to notice is a non-voluntary response that cannot be traced back to any prior culpable action or omission on his part, but can be ascribed to an underlying morally objectionable evaluative attitude that gives rise to his failure to notice. Thus Ryland’s telling an offensive joke arises from her failure to notice that her listeners are insulted, and her failure to notice arises from her insufficient lack of concern for the well-being of anyone beyond herself. In view of our discussion, the theorist who holds that this is a non-tracing case of culpable ignorance must argue that Ryland is blameworthy for telling the joke because she is blameworthy for failing to notice her audience’s discomfort, and that she is blameworthy for failing to notice her audience’s discomfort because her failure to notice arises from morally
objectionable evaluative attitudes on her part. He must further argue that Ryland’s failure to notice arises, not just from an isolated objectionable attitude triggered automatically by some stimulus, but rather from a full enough set of her evaluative attitudes at the time that they adequately represent her moral personality. In such a case this view entails that she is blameworthy, not just criticizable, for her failure to notice. As I have described Ryland’s case, this may be a plausible claim. Thus it appears there may be some non-tracing cases in which the agent is blameworthy for her wrongful act because, ultimately, she is blameworthy for the evaluative attitudes that underlie her failure to notice.

However, there will be many other non-tracing cases in which it is much less plausible that the agent is to blame for the failure to notice and hence for the act. In some of these cases, as Matt King has persuasively argued, it is hard to discern any morally objectionable evaluative attitude on the part of the agent as what gives rise to her failure to notice. In *Hot Dog*, for example, it is hard to see what is problematic about any of Alessandra’s evaluative attitudes, since she cares greatly about the family dog. Of course, we can imagine that if her attitudes were different, she would not have forgotten the dog in the car. For example, if she were less concerned about her children’s welfare and more concerned about the dog’s welfare, perhaps she wouldn’t have forgotten the dog. But we can hardly say that such a configuration of attitudes would make up a morally better psychology.62

There are many other cases, however, in which the failure to notice arises from an objectionable evaluative attitude – but it arises from such a limited set of attitudes that it is inappropriate to hold the person blameworthy for this limited flaw. Cognitive states – noticing or not noticing – are like actions and emotional responses in that they can be triggered by environmental stimuli that activate a single evaluative attitude which gives rise to the cognitive state, but in a situation in which the agent’s other pertinent evaluative attitudes are not activated. The sudden appearance of a voluptuous blonde pedestrian may divert a driver’s attention and thereby render him momentarily unaware of a car changing lanes immediately in front of him. A dramatic moment on a television reality show can rivet a babysitter’s attention, and prevent her from noticing that the baby is crawling out of the room. A wife’s unexpected announcement that a neighbor has been fired from his job can distract a husband just as he is turning on the power saw. An overheard conversation at a professional conference can break into an attendee’s consciousness, and make her lose the thread of what her companion is saying.

In each of these cases the agent’s failure to notice may result in an unfortunate action, and in each case we may want to say the agent should have noticed. We may condemn the evaluative attitude that gave rise to the failure to notice, and think worse of each agent for having such an attitude. (It is one thing for a driver’s attention to be

---

62 This suggests some of the difficulties with spelling out what a “reprehensible” or “morally good” configuration of desires and aversions is. Does it have to do with the upshots of the configuration on that very occasion, or with the upshots of that configuration over some longer span of the agent’s life? Or does it not have to do with “upshots” at all, but rather with the content of the configuration? This question mirrors some of the problems that arise in trying to construct a plausible form of rule utilitarianism.
diverted by a voluptuous blonde, and another thing for his attention to be diverted by the sound of a loud explosion, which might after all affect traffic conditions ahead.) But it is also true in each case that the evaluative attitude that gave rise to the failure to notice does not reflect anything like what would be the agent’s full evaluation of the situation. The driver is not, all things considered, willing to risk an accident in order to check out the passing blonde; the babysitter is not, all things considered, willing to risk harm to the baby in order to watch the denouement of the show; the husband is not, all things considered, willing to risk an injury in order to hear his wife’s news; and the conference attendee is not, all things considered, willing to insult her companion in order to eavesdrop on the other conversation. Thus, on the account for which I am arguing, we have no grounds for holding these agents blameworthy for their wrongful acts, since these acts issue from failures to notice that in their turn do not reflect a sufficient component of the agent’s full set of evaluative attitudes to show that he or she is blameworthy for the non-voluntary response. Each of these agents fails to notice something important; the failure to notice leads to a wrongful act; and the failure may arise from an undesirable evaluative attitude (the driver’s disproportionate interest in a sexually attractive stranger, the babysitter’s over-involvement in the reality show, the husband’s impulse to gloat at his neighbor’s misfortune, the conference attendee’s fascination with the opinions of her professional superiors). But in each case an environmental stimulus compels the agent’s attention in such a way that his or her other evaluative attitudes, which might have led to a different course of action, remain inactive. Hence we cannot judge that the incident reveals the agent’s full moral personality to be reprehensible.

If this is correct, in many cases we can appropriately think worse of a person who displays a non-voluntary response, such as a failure to notice, that reflects an underlying morally objectionable evaluative attitude. But we cannot hold her responsible for it in a way that renders her blameworthy, since it only indicates part of her evaluation of this situation that triggers it. To hold her blameworthy, we need to know the full story, and her non-voluntary response does not tell us that. Her response reveals part of her moral personality, but it does not reveal enough of her moral personality to render her culpable.

Ascertaining which cases reflect too little of the agent’s moral psychology to render her blameworthy, which ones reflect enough to render her somewhat blameworthy, and which ones reflect enough to render her fully blameworthy, is not an easy practical task. If I am right about the requirements for blameworthiness, we are faced with a continuum of cases in which there are degrees of blameworthiness, and the degree of blameworthiness depends on aspects of the case of the kind that are rarely articulated in philosopher’s examples, and that may be extremely hard to ascertain in real life. I suspect our actual judgments, which often pull us in several different directions, already reflect sensitivity to the importance of these considerations.

---

63 Even George Sher (Sher 2009, p. 25) points out that “We all know what it is to be assaulted by an urgent problem that drives all other thoughts from our minds….”

64 Thanks to Douglas Husak for pointing out the infrequency with which philosophers’ examples provide all the necessary information. For an account of differential reactions to these kinds of cases, see, for example, Weingarten (2009).
G. Conclusion

Many philosophers have argued that there are cases in which an agent is blameworthy for an act that is performed in ignorance, when the ignorance is itself blameworthy. It is argued that some of these cases can be understood as “non-tracing” cases, in which the agent’s culpability for her ignorance does not trace back to any earlier culpable failing on the agent’s part. Instead, it is claimed by Attributionists, the ignorance is culpable because it arises from a morally reprehensible evaluative attitude on the agent’s part that immediately gives rise to her failure to notice. Philosophers urging this perspective have typically agreed that the agent’s failure to notice is a non-voluntary response, but have maintained that we can be held blameworthy for such non-voluntary responses, just as we can be held blameworthy for our voluntary acts.

In this paper I have outlined a fresh argument that the differences between voluntary acts and non-voluntary responses are not sufficient to show that we can never be blameworthy for our non-voluntary responses. I then argued that to ascertain the conditions under which a person can be blameworthy for her non-voluntary responses requires a deeper investigation of the conditions under she can be blameworthy for her voluntary actions or decisions. Blameworthiness for a decision to perform an action, I argued, arises in either of two circumstances: either when the agent believes that the act is all-things-considered wrong and chooses it anyway; or when the agent’s desires and aversions that give rise to the decision are such that, if the action had exactly the features which gave rise to the agent’s activating desires and aversions, then the action would be all-things-considered wrong. But there is a further caveat on this: the desires and aversions that give rise to the decision must represent a sufficient portion of the agent’s relevant desires and aversions that her decision reveals something close to her full moral personality at work. An agent can be judged to have a bad motive, but for her decision to qualify as blameworthy, it needs to sufficiently reflect her full array of relevant motives.

There are many cases in which an act arises from a morally objectionable desire. We can view such an agent as morally flawed insofar as she has this desire. But if the desire is automatically activated by an environmental stimulus in circumstances in which no other of the agent’s desires and aversions are recruited to participate in the decision to perform the action, then the decision fails to reflect enough of the agent’s moral psychology to underwrite a judgment that she is blameworthy overall for what she does or decides to do.

Turning back to the question of non-tracing failures to notice, there may well be cases in which the agent’s failure to notice arises from underlying morally reprehensible evaluative attitudes, and reflects a sufficiently representative swath of the agent’s attitudes that it is legitimate to blame her for her non-voluntary failure to notice. But there are many other cases in which an agent’s failure to notice arises from a morally reprehensible evaluative attitude which in turn is triggered by an environmental stimulus that only recruits a single attitude, not the agent’s full set of relevant attitudes. In such cases we can fault the agent for her reprehensible attitude, but I have argued that it is
inappropriate for us to judge her to be blameworthy for the failure to notice or the ensuing wrongful act, since the failure to notice only reflects a small portion of her psychology, not a sufficient portion of it for us to judge her to be overall morally reprehensible for what she has done – the judgment that is required when we appraise someone as blameworthy. There is ground for moral criticism of such an agent, but no ground to condemn her as blameworthy for her wrongful act.
Acknowledgements

For comments on earlier versions of this paper I am grateful to Alvin Goldman, Karen Shanton, discussants at Union College, and especially Douglas Husak and Kenneth Simons.
References


Witmer, J. (2010, August 13). Interview on All things considered. National Public Radio Station WNYC.
