33. See N. Badhwar (2008) "Is Realism Really Bad For You? A Realistic Response," *The Journal of Philosophy* 105 (2), 85–107.
34. This paper was presented at Vreije Universiteit, Amsterdam, on May 7, 2013. I am grateful to the audience for helpful questions and comments. I also thank Mike Austin for inviting me to contribute to this volume.

11

Virtue Ethics and Moral Failure: Lessons from Neuroscientific Moral Psychology

Lisa Tessman

I Introduction

Recent empirical work in moral psychology has shown that moral judgments, like many other kinds of judgments, arise from two (somewhat) distinct systems: an automatic intuitive system that produces most of our moral judgments, and a controlled reasoning system that can be, though usually is not, engaged in the production or revision of moral judgments. This chapter is premised on the assumption that being a morally good person requires engaging both of these neural systems. I situate this assumption within a loosely Aristotelian virtue ethics framework, where being a good or virtuous person requires both reasoning and the habituation of virtues; when virtues are successfully habituated, the virtuous person is able to respond automatically in morally praiseworthy ways.

The dual-process model of moral judgment sheds light on the phenomenon that I will focus on here – namely unavoidable moral failure[1] – while rationalist models tend to obscure this phenomenon. Both deontology and certain forms of consequentialism, for instance, recognize only the reasoning process for arriving at a (justified) moral judgment, and because the reasoning process can eliminate impossible moral requirements either (for deontology) through a logical exercise with "ought implies can" as a premise or (for consequentialism) through a cost–benefit analysis that admits only possible options for consideration, neither of these moral frameworks can countenance the phenomenon of unavoidable moral failure. One might think that virtue ethics cannot countenance it either. However, if virtue requires making automatic intuitive judgments as well as reasoned moral judgments, and if the intuitive system – even, or especially, when it is functioning

excellently – can lead one to judge that one must, morally, do something that it is not possible to do, then virtue ethics *will* be able to make sense of the phenomenon of unavoidable moral failure. Hence in this chapter I hope to illustrate how (good) automatic, intuitive processing can lead a morally good person to the experience of inevitable moral failure.

Usually, automatic intuition and controlled reasoning work smoothly together. But given that intuitive and reasoned processes can be triggered by different stimuli, are underwritten by different kinds of affective responses, and involve different neural mechanisms, there is no reason to expect that an excellent controlled reasoning process and an excellent automatic, intuitive process would always yield the same verdicts. When the verdicts conflict and recommend actions that cannot both be performed, the morally good person faces a dilemma and is unable to carry out at least one of the actions. What I am suggesting is that the harmony between affect-laden intuition and reasoning – a harmony on which virtue has been premised – may often be impossible, or may be achieved only by sacrificing the excellence of either moral intuition or moral reasoning, and along with it, perhaps also sacrificing values that can only be sustained through one or the other process. Applying virtue ethics to dilemmatic situations reveals the inevitability of moral loss or failure.

II The dual-process model of moral judgment

Cognitive psychologists widely recognize two psychological systems for cognitive processing. There is "System 1," which "operates automatically and quickly, with little or no effort and no sense of voluntary control," and "System 2," which "allocates attention to the effortful mental activities that demand it, including complex computation" (Kahneman 2011, 20–21). System 1 is unconscious, associative, automatic rather than controlled, and fast. In contrast, System 2 is conscious, inferential, controlled, and relatively slow, and it takes effort to put System 2 to work. The operations of each system engage a number of different brain regions. The two systems can be brought into conflict, and the conflict itself activates another area of the brain.[2]

A variety of scientific methods have been used to investigate specifically *moral* cognition, yielding a dual-process model of moral judgment. The key claims of this model are that:

1. both an automatic intuitive system and a controlled reasoning system can take part in producing moral judgments, but they play different roles;

2. the intuitive system is "supported by affective processing" (Cushman, Young and Greene 2010, 57), though reasoning, too, depends on emotions of a different kind;
3. most moral judgments are made intuitively; and
4. judgments produced by the two processes can come into conflict with each other.

The first of these claims is built on the finding that moral cognition parallels other kinds of cognition in being carried out by both of these dual processes. Cognitive scientists arrived at the second claim by bringing together work on automaticity, which explains how intuitions are produced quickly and unconsciously, with work on the role of affect or emotion in moral judgment, and showing that in the case of moral judgment, intuitions are affect-laden in a particular way (Cushman, Young and Greene 2010). A moral intuition is "the sudden appearance in consciousness of a moral judgment, including an affective valence (good–bad, like–dislike), without any conscious awareness of having gone through steps of searching, weighing evidence, or inferring a conclusion" (Haidt 2001, 818).[3] The third claim, namely that most moral judgments are arrived at through the affect-driven, automatic intuitive process, is well supported and agreed upon even by those who disagree about how extensive a role is played by the controlled reasoning process.[4] Because intuitions are heavily affect-laden (claim two) and most moral judgments are made intuitively (claim three), "moral action covaries with moral emotion more than with moral reasoning" (Haidt 2001, 823). The fourth important claim of the dual-process model of moral judgment is that the dual processes can influence and respond to each other, but that the verdicts of each cognitive system may also come into direct conflict with each other; thus one can say that "moral judgment is the product of interaction and competition between distinct psychological systems" (Cushman, Young and Greene 2010, 47). I will unpack these claims below, with the aim of understanding the empirical details of what happens cognitively when one encounters a moral dilemma, and why such a situation can create an experience of inevitable moral failure.[5]

Jonathan Haidt puts the claim that most moral judgments are made intuitively at the center of his "social intuitionist" model. According to this model, most moral judgments follow directly from a subject's affect-laden intuitive response to a situation, and then reasoning takes place *post hoc* for the purpose of justifying the judgment to others; thus reasoning rarely actually causes or produces the moral judgment.

Haidt (2001, 2012) offers the metaphor of a lawyer defending a client to illustrate how *post hoc* reasoning defends intuitively produced judgments. The fact that reasoning usually plays the role of justifying prior intuitive judgments explains why intuition and reasoning are usually in agreement. However, people who form moral judgments in this way are not aware that they are doing so; they tend to believe that they have reasoned their way to the judgment. Additional evidence for the claim that conscious reasoning usually occurs after rather than before a moral judgment is made can be found in the phenomenon of what Haidt calls "moral dumbfounding" (Haidt 2001, 2012). Moral dumbfounding takes place when a subject makes a judgment (for example, that it is wrong to eat one's dead pet dog) and then is unable to come up with a reason to support the judgment, but nevertheless remains unshaken in her/his commitment to the judgment. Had conscious reasoning been what led the subject to the judgment in the first place, such moral dumbfounding would not take place – that is, the reason would still be readily available to the subject.[6]

While according to the social intuitionist model of moral judgment most moral judgments are made intuitively, reason does still have an important role: a *social* role. When one person gives reasons or arguments (that were formed *post hoc*) in support of a moral judgment, those reasons can affect *other* people, primarily by giving rise to intuitions in them (Haidt 2001, 819). A person's *own* reasoning is almost always "motivated" or biased in favor of supporting her/his own prior intuitive judgments, so private reasoning rarely brings about a change in one's own moral judgments. However, another person's reasoning – supporting an opposed judgment – is much more likely to change one's judgments. People are also often affected by others' judgments even when supporting reasons for the judgment are not supplied (Haidt 2001, 819). Reasoning that is *not* social or interpersonal and that changes one's own intuitive judgments is rare, but does exist. Haidt posits two ways in which one might reason one's way to a new moral judgment. The first way directly links reasoning with moral judgment: "people may at times reason their way to a judgment by sheer force of logic, overriding their initial intuition.... However, such reasoning is hypothesized to be rare, occurring primarily in cases in which the initial intuition is weak and processing capacity is high" (Haidt 2001, 819). This form of reasoning accounts for those times when intuitive and reasoned judgments come into conflict with each other. The second way indirectly links reasoning to moral judgment; the reasoning – for instance, reasoning in which one imagines oneself in someone else's shoes – triggers "a new intuition that contradicts the initial intuitive judgment" (Haidt 2001, 819) and then this conflict of intuitions must be resolved to produce the final moral judgment.

While Haidt's social intuitionist model emphasizes the emotionally infused intuitive process and locates reasoning's primary role in interpersonal communication, research by Joshua Greene and others has focused on identifying those cases in which people do use reasoning rather than an intuitive process to arrive at moral judgments. Greene's use of *f*MRI to study the neural processes that produce moral judgment have revealed that different moral situations (some more personal than others) tend to set different processes in motion (some affective and intuitive, others reasoned), and to lead to different judgments; his general finding is that "some moral dilemmas... engage emotional processing to a greater extent than others..., and these differences in emotional engagement affect people's judgments" (Greene et al. 2001).[7] When reasoning is used to independently produce a moral judgment rather than to defend a prior intuitive moral judgment, the reasoned judgment and the intuitive judgment may conflict.

Greene and colleagues tested subjects by giving them prompts of moral dilemmas that are paradigmatic in philosophical ethics because they bring deontological requirements into conflict with consequentialist considerations: a deontological prohibition makes one action forbidden, while that same action is prescribed by a consequentialist because it maximizes some good. These dilemmas include the variants that together comprise the "trolley problem."[8] For instance, the "Switch" dilemma goes like this: *An empty, runaway trolley is headed down a track on which five people are trapped; by flipping a switch you can divert it onto a sidetrack on which only one person is trapped. Should you flip the switch if this is the only way to stop the trolley from running over the five people?* In the variation that I will call "Push,"[9] the dilemma changes to this: *An empty, runaway trolley is headed down a track on which five people are trapped; by pushing a heavy person off of a footbridge over the track into the path of the onrushing trolley, you can cause this person's body to stop the trolley before it reaches the five people, but the heavy person will be killed in the process (you yourself are too light to use your own body for the purpose). Should you push the heavy person if this is the only way to stop the trolley?* Most subjects judge that it is appropriate to take actions relevantly similar to flipping the switch, but not to take actions relevantly similar to pushing the large stranger from the footbridge; that is, they behave like consequentialists in Switch (maximizing the lives saved, but violating a prohibition against killing) but like deontologists in Push (complying with

the prohibition against killing, while failing to maximize lives saved). Greene et al. found that in dilemmas that are like Push there is much more activity in the brain areas associated with emotion than there is in dilemmas that are more like Switch.[10] There is a correlation between subjects' experiencing a strong, negative emotional response (as most of them do in cases like Push) and subjects' judging it "inappropriate" (a stand-in for "morally wrong") to take the action (again, as most of them do in cases like Push). By also measuring response time, Greene et al. determined that respondents who judge it *appropriate* to take actions like pushing the large stranger first experience a strong, negative emotional response, and then take additional time to arrive at a judgment, time in which the brain can engage in controlled reasoning (e.g. weighing the costs and benefits of each action, and deciding that pushing has net benefits), can then detect and handle the conflict between the emotionally driven response and the opposed reasoned conclusion, and can ultimately exert cognitive control and override the emotional response (Greene et al. 2001). Further research (Greene et al. 2004) found that brain areas associated with abstract reasoning and cognitive control are more active during these longer response times.

By introducing moral dilemmas where the judgments in favor of each response are more evenly split than they are, for instance, in Push (where the vast majority of respondents judge it wrong to push), Greene and colleagues were able to compare brain activity in respondents who make opposite judgments. These dilemmas include those like the "Crying Baby" dilemma, which goes like this: *You and several others are hiding from enemy soldiers when your baby starts to cry; if the baby is allowed to cry the noise will alert the enemies, who will kill all of you, including your baby. Should you smother your baby if this is the only way to silence him/her and avoid alerting the enemies?* The emotional response – a powerful negative affective response to the thought of smothering one's baby – competes with the reasoned judgment that there is no benefit to refraining from smothering (the baby will still die). Greene and colleagues found that the brain areas associated with reasoning, with conflict, and with cognitive control are more active in subjects who give a verdict that it is appropriate to smother the baby than in those who give the opposite verdict (Greene et al. 2004). Later experimentation involved manipulation of neural processes by placing subjects under cognitive load and thus interfering with reasoning and cognitive control. In subjects who approve of smothering the crying baby, being under cognitive load slows response time, but in subjects who disapprove, there is no effect on response time, thus suggesting that it is reasoning – which is

affected by cognitive load because attentional resources for controlled processes are limited – that leads to an ultimate judgment of approval, and an emotional process – which is unaffected by cognitive load – that leads to judgments of disapproval. These findings are consistent with other research aimed at showing a causal relationship between the kind of cognitive process that takes place and the moral judgment that is produced. For instance, subjects with emotional deficits (due to ventromedial profrontal lesions) are more likely than healthy subjects to judge that it is appropriate to engage in actions that healthy subjects tend to find strongly aversive, like pushing the person off the footbridge; the absence of strong negative affect allows the reasoning process to dominate the judgment, and this leads to approval of pushing (Koenigs et al. 2007). In other experiments, subjects who have been manipulated to have positive emotions (by being shown a funny video clip) that serve to counterbalance otherwise negative emotional responses (such as to the thought of pushing) also are more likely than control subjects to approve of actions like pushing the large person to his death (Valdesolo and DeSteno 2006).[11]

To understand how making an intuitive moral judgment *feels* different than making a moral judgment on the basis of a consequentialist process such as cost–benefit analysis, it is necessary to examine the role of affect in each experience; so far I have portrayed the intuitive process as affect-driven, but have given no details about the role of affect in the reasoning process. There is evidence that rules or principles that are applied or manipulated through reasoning are themselves originally dependent on emotional responses, and so emotions are crucial to *all* moral cognition, even the operations of the reasoning system.[12] If something on which a moral principle depends conceptually, such as harm, were affectively neutral, there would be no motivation to avoid it and so no motivation to formulate or follow a moral principle that guides one to avoid it. The motivation to avoid harm comes from an affective experience (and similarly with moral concepts other than harm). Thus in considering the origin of a utilitarian principle, Cushman, Young and Greene (2010) propose that "affect supplies the primary motivation to regard harm as a bad thing," and then a controlled reasoning process "uses this core statement of value to construct the utilitarian maximum [*sic*] that we ought to act so as to minimize harm" (62). If this characterization is correct, then affect plays a role in both the intuitive process and the reasoning process, but it is a different role. What I have been calling the "emotional" or "affect-laden" intuitive process is really a process that involves emotion or affect in a particular way.

Greene's (2008) explanation for how affect can play a different role in intuitive and (utilitarian) reasoned moral judgments is that there are (at least) two basic kinds of emotional responses; the kind that plays a role in the intuitive process is different than the kind that plays a role in the cognitive process of reasoning that yields consequentialist judgments. Greene offers a metaphor for these two kinds of emotional responses: the kind that gives rise to (at least some) intuitive moral judgments are like *alarm bells*, while the kind that determine the values and disvalues that can be traded off in consequentialist reasoning are like *currency*. These two kinds of emotions function differently:

> Alarm-bell emotions are designed to circumvent reasoning, providing absolute demands and constraints on behavior, while currency emotions are designed to participate in the process of practical reasoning, providing negotiable motivations for and against different behaviors... Currency-like emotions function by adding a limited measure of motivational weight to a behavioral alternative, where this weighting is designed to be integrated with other weightings in order to produce a response. (Cushman, Young and Greene 2010, 62–63)

Alarm-bell emotions issue non-negotiable commands – "'Don't do it!' or 'Must do it!'" (Greene 2008, 64) – that (if not intervened with) automatically trigger some particular action. These commands "can be overridden," but "are designed to dominate the decision rather than merely influence it" (Greene 2008, 64–65). In contrast, currency emotions tell one what is valuable, and how valuable, so that they can influence a decision, but only in proportion to their value; that is, they are well suited for being weighed, and potentially *out*weighed. They offer information like "'Such-and-such matters this much. Factor it in'" (Greene 2008, 64); this information cannot be turned into an action-guiding decision until the weighing or calculating process has taken place.[13]

While it is primarily consequentialist and deontological frameworks that have been invoked in discussions of the philosophical implications of neuroscientific findings in moral psychology, there are also important implications for virtue ethics.[14] My assumption, from within a virtue ethics framework, is that a morally good person must have excellent practical reasoning and must also be habituated to respond automatically in morally praiseworthy ways; the dual-process model supplies the empirical details of how both responses take place. Part of what the empirical data have made clear, however, is that the reasoned and affect-laden automatic responses can diverge, and that given certain dilemmatic situations, they typically will diverge. Deontologists and consequentialists may not be ready to dispense with the automatic process altogether, but they do characterize these responses as *ir*rational, and emphasize that they cannot justify – i.e. provide a reason for – an action. While deontologists and consequentialists typically differ on what they take the correct reasoned judgment to be, they agree that judgments arising from unconscious, automatic processes are not to count as justified moral judgments unless a reasoned justification is also given to support the judgment. This allows them to make normative claims about how to resolve conflicts between automatic and reasoned responses: reasoning can and must always justify the action-guiding decision in cases of conflict, and this decision fully resolves the conflict, given the assumption that an overridden automatic response, having been shown to be unjustified, completely loses its normative force. I will suggest that virtue ethics should not follow suit in the dismissal of *a*rational processes of moral judgment, even though this means, among other things, admitting that conflicts cannot always be resolved without "remainder"[15] and thus that there are situations of unavoidable moral failure. That is, conflict situations are one kind of situation in which one may find that one morally must do what it is impossible to do. These situations have been discussed extensively in the "moral dilemmas debate," and I will not repeat those discussions here;[16] instead, I will present an empirical explanation of dilemmatic conflicts, and then consider its implications for virtue ethics.

Empirical work shows that one person can respond both automatically and through controlled reasoning to the same situation, with the two processes generating opposite judgments; when this happens one is internally conflicted and might remain conflicted in some way, even though for the purpose of choosing how to act one must eventually arrive at a single action-guiding decision. Fiery Cushman and Liane Young take the experience of moral conflict to be a direct consequence of the fact that "a number of distinct psychological mechanisms accomplish moral judgment in ordinary people," noting that "these mechanisms sometimes conflict within a single individual, giving rise to the experience of a moral dilemma" (Cushman and Young 2009, 10). Dilemmas like Crying Baby reliably evoke a psychologically conflicted response (in most subjects), because both the automatic intuitive response ("Don't hurt the baby!") and the reasoned response ("The baby will die either way, so I must choose between saving the lives of everyone else or saving no lives, and saving some lives is better than saving no lives, thus...")

are very compelling (whereas in both Switch and Push, most people find only one of the two possible options to be very compelling – the reasoned response in the case of Switch, and the intuitive response in the case of Push). In cases like Crying Baby, neuroscientific studies indicate brain activity corresponding to both processes taking place, and "reveal signatures of cognitive conflict: a neuronal reconciliation between the competing demands of separate psychological mechanisms" (Cushman and Young 2009, 11). Cushman and Young propose that, to represent the internal conflict that people experience when their different psychological processes produce clashing judgments about a case, people could deliver the verdict that a case is a "dilemma" rather than be restricted to the judgment that an action is either "forbidden, permissible, obligatory, supererogatory and so forth" (Cushman and Young 2009, 17). In dilemmas like Crying Baby, resolving the conflict in favor of either judgment for the purpose of action-guidance does not resolve the psychological conflict; thus a dilemma is marked by its distinctive *psychological* feature: "When you face a dilemma, no matter what you do, part of you is going to be dissatisfied" (Cushman and Young 2009, 19). If engaging in multiple kinds of cognitive processing is inescapable for (most) humans and if the outputs of each process will sometimes conflict, then moral life will necessarily be experienced as dilemmatic.

Furthermore, the experience of conflict between two moral judgments does not arise *only* from dual processing; it can arise from intuitive processing alone. Verdicts produced by reasoning alone cannot be dilemmatic, for the reasoning process can and must eliminate all conflicts, either logically (for instance, utilizing the principles of deontic logic) or through calculations that absorb costs into benefits (thus fully resolving conflicts between *prima facie* moral requirements and yielding one all-things-considered moral requirement). However, the intuitive process can produce dilemmatic moral judgments, for two alarm-bell emotions could command incompatible actions; *Sophie's Choice* is a classic example of this (see Styron 1976). If an alarm-bell emotion gives rise to an intuitive judgment that an act morally must be performed, then whether it clashes with another intuitive judgment or with a reasoned judgment, it will remain standing even if overridden; there is nothing in the intuitive process that can eliminate it.[17] The "dissatisfaction" of the intuitive system may differ in kind from the "dissatisfaction" of the reasoning system, just as the emotions that give rise to intuitive and reasoned judgments differ in kind. That is, failure to heed an alarm bell will feel different from the failure to maximize the values that are experienced as being like currency.

III The dual-process model and virtue ethics

While virtue – in an Aristotelian framework – requires both right reason and right desire, and requires that right reason and right desire point in the same direction, the dual-process model indicates that there are situations where excellent reasoning and excellent affect-laden intuitions will point in different directions, thus creating a conflict; in these cases it would be wrong *not* to have both the automatic response and the reasoned judgment, but having them both – with a conflict between them – must be said to preclude virtue if one insists that virtue requires harmony between automatic, affective responses and reasoned responses. Virtue ethicists such as Rosalind Hursthouse who have theorized about moral dilemmas have noted that when faced with certain kinds of moral dilemmas, two different virtuous people may make different decisions about what is best – indeed, it is the mark of an irresolvable dilemma that this can happen (Hursthouse 1999, chapter 3). The neuroscientific work in moral psychology that I have been discussing allows the virtue ethicist to go one step further with this analysis: when faced with a dilemma, a single person may be pulled in two different directions precisely because, as found by Cushman and Young, her/his reasoning process and her/his automatic, intuitive process may deliver different verdicts, each verdict indicative of the excellent functioning of one cognitive system. The conflict of verdicts is not a symptom of a defect in either process; a person who engages in an excellent reasoning process and who has an excellent affect-laden intuitive response can experience such a conflict. Indeed, in some situations, the absence of psychological conflict would be indicative of a deficit in either the reasoned or the affect-laden intuitive response. For instance, in facing a situation in which both of two people are endangered but one cannot save both, if one were to lack the alarm-bell emotions ("must save!") that underlie the two conflicting intuitive responses, one would exhibit not an excellence, but rather an emotional deficit. In a case like Crying Baby where it is a reasoned and an intuitive response that conflict, one could plausibly claim that the best reasoned judgment is to stop the baby from crying, while the best intuitive judgment is the one backed by the "don't smother the baby!" alarm-bell emotion.

What should a virtue ethicist make of this? I believe that virtue ethicists should resist joining deontologists and consequentialists in their denial of the possibility of unavoidable moral failure brought on by the dilemmas that arise when intuitive moral judgments conflict with other intuitive moral judgments or with reasoned moral judgments. Deontologists and

consequentialists eliminate this possibility by dismissing intuitive moral judgments, that is, by refusing to give any normative weight to intuitive judgments. It is this move, I propose, that virtue ethicists should not make; that is, virtue ethicists should not conceive of the virtuous agent as solely a rational agent, and then simply determine what such an agent's reasoned moral judgment should be.

Some virtue ethicists, such as Nancy Snow, come close to doing this, not by denying that automatic or intuitive moral judgments are legitimate or by claiming that only reasoned moral judgments matter, but rather by envisioning the automatic responses of a virtuous agent to be always in line with that agent's reasoned judgment. Snow (2010) describes traditionally conceived virtues in terms offered by social psychologists, including those who work on automaticity. While Snow's aim is primarily to counter the claim that there are no traits that are sufficiently "global" to count as virtues, along the way she considers what virtues' relationship to rationality can be if virtues are incorporated into habituated traits. Snow, presumably in order to hold onto the traditional picture of virtues as requiring excellence in both reasoning and affect as well as a harmóny between them, sketches several ways in which virtues can still be said to be enacted "for a reason," even when they have become habitual and unconsciously prompt the moral agent to act. The first way is that one can deliberate specifically about one's habitual responses, and consciously work to change them to bring them more into line with one's reasoned beliefs. Snow refers to work in the psychology of prejudice – where there is evidence that a deliberate process can alter one's automatically activated stereotypes – as an example of how reasoned choices enter into the formation of new automatic or habitual responses (Snow 2010, chapter 1). The second way that Snow conceives of reasoning as informing virtues is through what she calls *goal-dependent automaticity*. Some automatic actions are done for a reason in the sense that they are done "to serve the agent's chronically accessible goals," where the goals themselves were formed through a reasoning process. In such cases, "the agent's reason for acting – to serve a chronic goal – is not present to her consciousness at the time of acting, but is operative in her psychological economy, and is such that, were it brought to her conscious awareness, she would endorse it as her reason for acting" (Snow 2010, 51). Goal-dependent automatic actions have become automatic because "repeated encounters with situational cues trigger an agent's virtue-relevant goals outside of her conscious awareness, resulting in her habitual performance of virtuous actions in those circumstances" (Snow 2010, 61).

Without disagreeing with Snow that virtues – including their affective components – *can* be dependent upon the agent having at some time reasoned her way to a goal that later comes to be pursued automatically, dual-process models – and especially Haidt's social intuitionist model – suggest that this is not what tends to happen. If Haidt is right about how rarely moral reasoning affects (one's own) moral judgments, then virtues are developed and enacted much more intuitively (and less through reasoning) than virtue ethicists such as Snow assume that they can be. In fact, Haidt and Joseph (2004) argue that "virtues... are closely connected to the intuitive system" (61) and point to affectively valenced "flashes" of intuition – which they argue are innate – as the "building blocks that make it easy for children to develop certain virtues and virtue concepts" (63). A virtue ethicist who wants to hold on to the rationality of the virtuous agent's judgments might reply with something like this: 'Haidt may be right that people's moral judgments are not *typically* dependent on reason, but that is precisely what makes the virtuous agent special or atypical; the virtuous agent, unlike most others, has managed to shape even her automatic moral judgments to fit rationally chosen goals.' I think that taking this tack would be a mistake, for it presupposes that there are no human responses that are both morally valuable (and thus necessary for full virtue) and arational, that is, independent of rational processes. I believe that there are moral values that are – and probably can only be – upheld through arational processes that do *not* rest on prior reasoning. If a moral agent lacked the automatic, intuitive responses that support these crucial values, it would be wrong to call that agent virtuous, no matter how well she/he deliberated, and no matter how well her/his goal-dependent automatic responses harmonized with her/his reasoned judgments.

Among those automatic intuitive moral responses that I believe do *not* rest on prior reasoning are those that have parallels in the automatic processes of other mammals, who lack the neural systems for the reasoning in which humans can engage. Patricia Churchland, for instance, argues that the human "neural platform" for morality is largely shared with other mammals (though goes beyond that of other mammals in important ways), and includes such things as the release of oxytocin, which enables trust and attachment, which in turn underlies the "alarm-bell" emotional responses that occur when, for instance, a loved one is threatened.[18] Someone's "don't smother the baby!" alarm-bell response in the case of the Crying Baby dilemma is best described as an entirely intuitive process, and would be misrepresented if it were described as serving the rationally chosen "chronic goal" of being a good parent. It is excellence

of affect-laden intuition – which is not dependent on any rationally chosen goal – that is displayed by this alarm-bell response. If excellent reasoning is conducted about the Crying Baby dilemma, it may very well deliver an opposed, rather than a harmonious, verdict.

Philip Tetlock offers another way to appreciate how an excellent intuitive moral judgment could lack harmony with an agent's moral reasoning. Tetlock and colleagues empirically document how the importance of some values has been marked by the fact that people have sacralized these values, and upholding these values can only be accomplished intuitively – reasoning *destroys* them and disqualifies the reasoner from the relationships whose core they form. A sacred value is defined as "any value that a moral community implicitly or explicitly treats as possessing infinite or transcendental significance that precludes comparisons, trade-offs, or indeed any other mingling with bounded or secular values" (Tetlock et al., 2000, 853). When values are sacralized there is a risk that they will be subjected to "taboo trade-offs" – namely trade-off comparisons of a sacred with a non-sacred value – or "tragic trade-offs" – trade-off comparisons between two sacred values. The research finds that indeed people do, psychologically, treat certain values as sacred, and certain trade-offs as either taboo or tragic. This is manifested by the fact that subjects express moral outrage about (fictional) decision-makers who merely contemplate taboo trade-offs (with greater outrage for those who choose to sacrifice a sacred value than for those who protect it), and when they themselves are pressed into considering taboo trade-offs, they demonstrate a desire to cleanse themselves morally afterwards by, for instance, supporting a cause like organ-donation (Tetlock 2003). The longer that a decision-maker spends contemplating – that is, reasoning about – a taboo trade-off, the more negatively observers will rate him or her. For example, in a narrative about a hospital administrator who must decide whether to spend funds to save the life of a child or to use the same funds "for other hospital needs," if subjects are told that the hospital administrator decides "after much time, thought, and contemplation" to save the child's life, they express intense moral outrage about him, but they do not if they are told that the administrator is very quick to make the decision to save the child's life (Tetlock et al., 2000, 858). In other words, thinking about the unthinkable is treated as a moral transgression, and the more one thinks, the worse it is: "Even when the hospital administrator ultimately affirmed life over money, his social identity was tarnished to the degree that observers believed that he lingered over that decision. It was as though participants reasoned 'anyone who thinks that long about the dollar value of a child's life is morally suspect'" (Tetlock et al., 2000, 856). On the other hand, if the narrative is altered so that the hospital administrator must choose to either save the life of one child or save the life of another child, thus leading subjects to treat the situation as requiring a tragic (rather than taboo) trade-off, then they praise the administrator for spending *more* time deliberating; when the sacrifice of a sacred value is inevitable, longer deliberation signals a deeper desire to prevent this inevitable sacrifice (Tetlock et al., 2000, 860; see also Fiske and Tetlock 1997, Tetlock 2003).

I take Tetlock and colleagues' research to show that, psychologically, people take some things to be appropriately valued only when moral judgments about them are made intuitively. For instance, the judgment that one must, morally, protect a human life can be a form of *de*valuing that life if the judgment is made through a cognitive process that is inappropriate for it. In cases like this, it is the very move from intuition to reasoning that constitutes a betrayal of values that are in part constituted by their guaranteed insulation from the negotiations that take place through conscious reasoning. The moral judgment to protect a sacred value must be made through an automatic process, and the verdict of this process could conflict with the same agent's reasoned moral judgment.

I believe that virtue ethicists would do well both to recognize the sort of values that are, and can only be, supported through automatic, intuitive moral responses, and to acknowledge the potential for conflicts amongst these responses, or between these responses and a reasoned judgment. Nevertheless, I am not arguing for the sacrifice of values that are best achieved through reasoning, such as the value of fair and impartial treatment of others. Rather, my claim is that, due to the fact that some values are achieved through reasoning and some through intuition, and given the impossibility of realizing them all, moral life is, through and through, dilemmatic. In such a condition, what constitutes virtue? What does it mean to "apply" virtue ethics to a dilemmatic moral life?

While it would be absurd to call someone virtuous who *lacked* crucial affect-laden automatic responses, it also seems that someone who experiences dilemmatic conflicts – that is, situations of unavoidable moral failure – due to these automatic responses would lack the requisite harmony to count as virtuous. The quest for virtue under dilemmatic conditions can only go wrong, it seems: we must squelch crucial alarm-bell emotions in order to avoid the possibility of their producing moral conflicts, but if we are to maintain our attachments and the

sacredness of the values arising from them, we must continue to pay attention to alarm-bell emotions and the intuitive moral judgments that they support. While an alarm-bell might have to be overridden for the purpose of action-guidance in a dilemma, doing so should *not* be facilitated by the agent ceasing to hear the alarm-bell, for this would indicate an emotional deficiency. The fact that dual processes can yield conflicting moral judgments does not mean that one of the processes should be curtailed; rather, it means that moral agents whose dual processes are both in good working order may experience conflicts rather than harmony. Furthermore, if one tries to expand the domain of one's alarm-bells so that one experiences them as requiring one to respond to more and more distant strangers, not only will the potential for conflict increase, but one may find that one is actually un*able* to have a sufficiently strong affect-laden response to so many people, and at such a distance; if this is the case, it just reveals that there may be an upper limit to (excellence in) human morality. When in situations that require something impossible – that require something past this limit – virtue becomes unattainable. Instead of forcing an "application" of an unattainable ideal, one will do better to focus on how best to survive and cope with the unavoidable moral failures that are supported by the kind of intuitive moral processing that humans do. One might call this a sort of non-ideal(ized) virtue,[19] but it is a far cry from the harmonious fitting together of affect-laden intuitions and controlled reasoning.

Notes

1. Here I borrow from C. Gowans (1994) *Innocence Lost: An Examination of Inescapable Moral Wrongdoing* (Oxford: Oxford University Press), who speaks of "inescapable moral wrongdoing."
2. The relevant brain areas are discussed and illustrated in J. Greene and J. Haidt (2002) "How (and Where) Does Moral Judgment Work?" *TRENDS in Cognitive Sciences* 6: 517–523.
3. There are different ways in which affect could be connected to moral intuitions. One plausible model describes intuitions as a form of heuristics and then posits an "affect heuristic." See W. Sinnott-Armstrong, Walter, L. Young, and F. Cushman (2010) "Moral Intuitions," in John Doris (ed.) *The Moral Psychology Handbook* (Oxford: Oxford University Press), 246–271.
4. For instance, Joshua Greene and Jonathan Haidt, who disagree about how often, and in what way, reasoning matters (and ought to matter) for moral judgment, agree that "emotions and reasoning both matter [for moral judgment], but automatic emotional processes tend to dominate" (Greene and Haidt "How (and Where) Does Moral Judgment Work?", 517).
5. In my explanation of the dual process model of moral judgment, I will draw primarily on the work of Jonathan Haidt and the work of Joshua Greene (as well as on the work of his colleagues, including Fiery Cushman and Liane Young); Haidt's and Greene's normative claims are at odds with each other, but they largely agree on their descriptive accounts of how moral judgments are made. See J. Haidt and S. Kesebir (2010) "Morality" in S. Fiske, D. Gilbert and G. Lindzey (eds.) *Handbook of Social Psychology*, 5th edition (Hobeken, NJ: Wiley), 807 for Haidt's remarks on the differences between his dual process model (which he calls the Social Intuitionist Model) and Greene's dual process model.
6. There is much additional evidence to support the hypothesis that reasoning takes place post hoc to rationalize intuitive moral judgments; see J. Haidt (2001) "The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment," *Psychological Review* 108: 814–834, and Haidt and Kesebir, "Morality." H. Mercier (2011) "What Good is Moral Reasoning?" *Mind & Society* 10 (2): 131–148 also presents evidence for the claim that moral reasoning does not serve the purpose of seeking truth, but rather serves the purpose of finding reasons and constructing arguments in support of one's own prior judgment, primarily in order to better persuade others, and evaluating the arguments that others try to use persuasively.
7. While this is consistent with Haidt's model – because Haidt does recognize that reasoning takes place in situations where the emotional influence is low – Haidt believes it is rare in ordinary moral life; if it is not rare in the experimental situations, this could indicate that Greene and colleagues' testing situations artificially prompt more reasoning than normally occurs in everyday situations. Haidt suggests that psychological interviews such as those conducted by Lawrence Kohlberg (whose rationalist model of moral development has been highly influential) do exactly this.
8. The first trolley (or "tram") case appeared in P. Foot (1978) "The Problem of Abortion and the Doctrine of the Double Effect" in *Virtues and Vices and Other Essays in Moral Philosophy* (Oxford: Blackwell), 19–32, and then developed into the "trolley problem" in J. Thomson (1985) "The Trolley Problem" *The Yale Law Journal* 94: 1395–1415.
9. I call it "Push" instead of "Footbridge" (its more typical label) because it is the thought of pushing rather than the thought of standing on a footbridge that elicits the emotional response that is the defining feature of the case.
10. Greene et al. classified dilemmas that are like "Push," as "personal, moral" dilemmas, as contrasted with dilemmas that are like "Switch," which were classified as "impersonal, moral" dilemmas (J. Greene et al. (2001) "An *f*MRI Investigation of Emotional Engagement in Moral Judgment," *Science* 293 (5537): 2105–2108). In later work (Greene et al. (2009) "Pushing Moral Buttons: the interaction between personal force and intention in moral judgment" *Cognition* 111: 364–371) Greene and colleagues revised the classification to more accurately capture the salient differences.
11. Greene's moral dilemmas are all designed to bring typically deontological judgments into conflict with typically utilitarian judgments. Although in all of his dilemmas that elicit strong emotional responses, it is the deontological judgment that is consistent with these emotional responses and the utilitarian judgment that requires overcoming the emotional responses, he still takes himself to have shown that typically deontological judgments are produced through an emotional, intuitive process (which is followed by *post*

hoc rationalizations produced through the reasoning process), and typically utilitarian judgments are produced through a controlled reasoning process (J. Greene et al. (2008) "Cognitive Load Selectively Interferes with Utilitarian Moral Judgment" *Cognition* 107: 1144–1154; see also J. Greene (2008) "The Secret Joke of Kant's Soul," in W. Sinnott-Armstrong (ed.) *Moral Psychology Vol. 3: The Neuroscience of Morality* (Cambridge, MA: MIT Press), 35–79). This conclusion has been challenged by G. Kahane et al. (2011) "The Neural Basis of Intuitive and Counterintuitive Moral Judgment" *Social, Cognitive, and Affective Neuroscience* (advance access published March 18, 2011): SCAN 1–10, who designed experiments using "scenarios where the utilitarian option is intuitive...and scenarios where the deontological judgment is intuitive," thus allowing them to "study the differential effect of the content (deontological/utilitarian) and the intuitiveness (intuitive/counterintuitive)" (p. 2). They found that "behavioural and neural differences in responses to [the dilemmas used] are largely due to differences in intuitiveness, not to general differences between utilitarian and deontological judgment" (p. 9). Thus it seems that *one* kind – but not the only kind – of reasoning that can overrule an emotionally driven intuitive response is the calculation or weighing of costs and benefits associated with the consequentialist rule of maximizing (expected) net benefits. (It is worth noting that F. Cushman, L. Young, and J. Greene (2010) "Multi-Systems Moral Psychology" in John Doris (ed.) *The Moral Psychology Handbook* (Oxford: Oxford University Press) acknowledge on pp. 60–61 that conscious reasoning can be nonconsequentialist and can take the form of applying deontological principles.) This consequentialist sort of reasoning is in fact the sort of reasoning that I would like to focus on when I ask what the dual-process model can illuminate about moral experience: what I am interested in is the fact that moral requirements grasped through affect-laden intuitions are experienced quite differently than moral requirements that are supported by consequentialist reasoning (such as cost-benefit analysis), so those experiments such as Greene's, where the conflict is between an emotionally driven judgment and a judgment produced by cost-benefit analysis, will illustrate this phenomenon well.

12. In other words, Hume was right.
13. While Greene's characterization of emotions as coming in two types – alarm-bells and currency – is somewhat speculative, there is evidence to support his hypothesis, cited in Cushman, Young and Greene, "Multi-Systems Moral Psychology."
14. As William Casebeer argues, "the moral psychology required by virtue theory is the most neurobiologically plausible" (W. Casebeer (2003) "Moral Cognition and its Neural Constituents," *Nature Reviews: Neuroscience* 4: 841–846, at 841), and as Haidt and Craig Joseph put it, "virtue theories are the most psychologically sound approach to morality. Such theories fit more neatly with what we know about moral development, judgment, and behavior than do theories that focus on moral reasoning or on the acceptance of high-level moral principles such as justice" (J. Haidt and C. Joseph (2004) "Intuitive Ethics: How Innately Prepared Intuitions Generate Culturally Variable Virtues" *Daedalus* 133: 62). For more discussion of the fit between virtue ethics and a dual processing model of moral psychology, see also J. Haidt and F. Bjorklund (2008) "Social intuitionists answer six questions about moral psychology" in W. Sinnott-Armstrong (ed.) *Moral Psychology, Vol. 2: The Cognitive Science of Morality: Intuition and Diversity* (Cambridge, MA: MIT Press),181–217. Casebeer assumes that "virtue theorists focus on the appropriate coordination of properly functioning cognitive sub-entities" and that "moral reasoning and action are therefore 'whole-psychology, whole-brain' affairs" (Casebeer "Moral Cognition and its Neural Constituents," 842). While it may be true that virtue ethics is the best framework for understanding the harmonious operation of "cognitive sub-entities" involved in moral judgment, I also think that virtue ethics offers a way to understand the moral implications of conflict – that is, *lack* of coordination or harmony – between different cognitive processes.
15. The idea of a moral "remainder" – which indicates that a moral requirement has *not* been eliminated and retains its normative force – comes from B. Williams (1973) "Ethical Consistency," in *Problems of the Self* (Cambridge: Cambridge University Press), 166–186.
16. On the "moral dilemmas debate" see, for instance: C. Gowans, ed. (1987) *Moral Dilemmas* (Oxford: Oxford University Press); C. Gowans (1994) *Innocence Lost: An Examination of Inescapable Moral Wrongdoing* (Oxford: Oxford University Press); H.E. Mason, ed. (1996) *Moral Dilemmas and Moral Theory* (Oxford: Oxford University Press); W. Sinnott-Armstrong (1988) *Moral Dilemmas* (Oxford: Blackwell Publishing); D. Statman (1995) *Moral Dilemmas* Value Inquiry Book Series 32 (Rodopi Bv Editions); M. Stocker (1990) *Plural and Conflicting Values* (Oxford: Oxford University Press) and Williams, *Problems of the Self*. See R. Hursthouse (1999) *On Virtue Ethics* (Oxford: Oxford University Press) for a discussion of the implications of the moral dilemmas debate for virtue ethics.
17. As Williams would put it, it cannot "eliminate from the scene the *ought* that is not acted upon" (Williams "Ethical Consistency," 175).
18. As Churchland writes: "The idea is that attachment, underwritten by the painfulness of separation and the pleasure of company, and managed by intricate neural circuitry and neurochemicals, is the neural platform for morality" (P. Churchland (2011) *Braintrust: What Neuroscience Tells Us About Morality* (Princeton, NJ: Princeton University Press), 16).
19. See L. Tessman (2009) "Feminist Eudaimonism: Eudaimonism as Non-Ideal Theory," in L. Tessman, ed., *Feminist Ethics and Social and Political Philosophy: Theorizing the Non-Ideal* (Dordrecht: Springer), 47–58; L. Tessman (2005) *Burdened Virtues: Virtue Ethics for Liberatory Struggles* (New York: Oxford University Press); and S. Gürsozlü (2010) *Virtues and Flourishing Under Oppression*. Ph.D. Dissertation, Binghamton University.