Moral thinking, more and less quickly

Mark Alfano\*, Joshua August Skorburg, Christina M. Karns,

Corresponding Author: Mark Alfano is Associate Professor at Delft University of Technology and Professor at Australian Catholic University. He uses tools and methods from philosophy and science to explore the field of moral psychology.

Delft University of Technology & Australian Catholic University

Jaffalaan 5, 2628 BX Delft,

Netherlands

mark.alfano@gmail.com

+31 638691845

Joshua August Skorburg is urrently ABD in Philosophy at the University of Oregon. Starting January 2018, he will be post-doctoral fellow in philosophy at Duke University. His research is in moral psychology and applied ethics.

University of Oregon Department of Philosophy

211 Susan Campbell Hall

Eugene, OR 97403.

g.skorburg@gmail.com

+1-815-979-7106

Christina M. Karns is a Research Associate scientist at the University of Oregon Department of Psychology. Her research uses human neuroimaging (fMRI and EEG), behavior, and assessments to characterize individual differences in emotion and cognition.

University of Oregon Department of Psychology

1227 University of Oregon

Eugene, OR 97403

+1-541-346-4261

Keywords: virtue, moral psychology, dual-process, experimental philosophy, attunement, fluency, timing

Acknowledgements: This work was supported by a grant from the Templeton Religion Trust to the University of Oklahoma Institute for the Study of Human Flourishing.

**Title:**

Moral thinking, more and less quickly

**Abstract (150 words):**

In this paper we argue that essential questions of empirical and moral psychology need a broader perspective that incorporates a virtue theoretic perspective. For example, dual-process systems are alluring in their ability to explain moral judgments in a simple framework in which a slow, often-overstretched rational system tends to produce consequentialist tendencies, while a fast, affective system produces virtuous (or vicious) tendencies that perform well in their habituated ecological niche but sometimes disastrously outside of it. However, difficulties with this approach are often overlooked. In this paper, we challenge some inherent assumptions of dual process models and aim to expand our moral theorizing beyond the clash between utilitarian reasoning and deontological intuitions by incorporating a virtue theoretic approach. We conclude by arguing that appropriate methods and approaches are available to address the shortcomings we identify and steer the field toward a more balanced view of a moral life.

Keywords: virtue, moral psychology, dual-process, experimental philosophy, attunement, fluency, timing

**1 Introduction**

What capacities and dispositions support human moral perception, judgment, emotion, motivation, deliberation, choice, and behavior? How and on what timescale are these capacities acquired and fine-tuned? How are they interrelated? How can they be measured? These are the essential questions of empirical moral psychology. Our aim in this paper is to establish some constraints on what acceptable answers should look like within the framework of *virtue ethics*, a family of philosophical theories that both furnish normative guidance about human flourishing and presuppose a suite of psychological structures and processes that enable people to pursue flourishing. If these structures and processes exist, they must do so in the bodies and brains of human animals as they interact with the world across different time scales. In principle, then, it should be possible to conceptualize these structures and processes precisely enough that they can be operationalized and measured in the lab and in the field to construct models of moral capacities. These models can then be used inform educational interventions to support moral development.

However, in empirically-informed philosophy and philosophically-sophisticated psychology, researchers have tended to ignore virtue ethics and focus on philosophical frameworks that lionize one or the other side of the dichotomy between fast, emotional, arational mental processes and slow, calculating, rational mental processes, emphasizing explicit moral decisions over intuitions (as pointed out in Narvaez & Lapsley, 2005). Following Hume (and, if Greene 2008 is right, Kant without Kant’s realizing it), one could valorize fast, hot, emotional moral mentalizing. Following Plato (and Kant, at least according to Kant), one could valorize slow, cold, rational moral mentalizing. But where does a mixed approach like virtue ethics fit into this scheme? For instance, in Aristotle, slower processes guide the shaping of habit, which in turn facilitates later, fast processes. Those fast processes, in turn, alert the virtuous agent to many features of the socio-moral environment (and blind her to others), prompting further slow processing and perhaps even laying down new habits, or reshaping or even eliminating old habits.

Although more recent dual-process theorizing allows for other interpretations of what fast and slow processes represent, by and large, these views continue to support a fairly rigid dichotomy. For example, Greene (2017) equates deontological cognition and decision-making with virtue-theoretic cognition and decision-making.[[1]](#footnote-2) We want to break away from this dichotomy by exploring the role played by a richer array of mental processes in the moral life and suggest ways that a virtue theoretic approach can contribute an alternative voice to debates in moral psychology.

Here is the plan for this paper. In Section 2, we articulate a multidimensional framework for moral psychological inquiry, focusing on the temporal domain. In Section 3, we spell out the virtue ethical constructs of attunement and fluent agency, arguing that these constructs commit the virtue ethicist to several *because*-claims that in turn entail temporal, spatial, and social constraints. In Section 4, we show that these constraints are not respected by research that employs a dichotomous distinction between slow and fast mental processes, or between hot and cold mental processes, and we argue that temporal claims would benefit from increased precision and pre-specification. We conclude with suggestions for how to move forward.

**2 A continuous multi-dimensional space for psychological inquiry**

Building on Christen & Alfano (2013), we can distinguish three of the main dimensions onto which moral psychological inquiry can be mapped: social, spatial, and temporal (Figure 1). The social dimension is discrete, indicating the number of people involved; the spatial and temporal dimensions are continuous. In this section, we describe the social and spatial domains and then elaborate on the temporal domain which is of primary relevance for our argument.



**Figure 1:** Three main dimensions onto which moral psychological inquiry can be mapped. The social dimension measures the number of people involved, whether as agents or patients (Gray, Waytz, & Young 2012), or both (Alfano 2016). The spatial dimension measures both the distance between people when more than one is involved, and the spatial scope of agential and decision-making processes within an individual. It is understandable though unfortunate that people are less likely to care about or help a statistical victim who is physically far away than one who is close (Singer 1972). The temporal dimension measures the amount of time that a given type of moral decision-making takes.

The social dimension measures the number of people involved, whether as agents or patients (Gray, Waytz, & Young 2012), or both (Alfano 2016). This dimension ranges from the single individual to dyads, triads, small groups, groups of approximately 150 (the “Dunbar number,” see Dunbar 1992, 1993), and even larger groups.[[2]](#footnote-3) Although the dimension is discrete and ordinal, the model does not assume that psychological processes map onto it linearly. As Alfano (2017) and Alfano, Huijts, & Roeser (forthcoming) have argued, the psychological mechanisms that serve us well when we are alone may not be as effective when we are in a close dyadic relationship (e.g., friendship, on which see Alfano 2016b and Alfano & Skorburg 2017). Likewise, the mechanisms suited to dyads may not be available or appropriate in larger social contexts — as the phenomenon of compassion fade suggests (Västfjall et al. 2014).

The spatial dimension measures both the distance between people when more than one is involved, and the spatial scope of agential and decision-making processes within an individual. It is understandable though unfortunate that people are less likely to care about or help a statistical victim who is physically far away than one who is close (Singer 1972). Again, psychological processes mapped to this dimension need not be linear. Moral decision-making can play out discursively between agents or individually within a particular agent. In addition, the spatial scope of the neural and somatic architecture that supports decision-making can vary or rely on dissociable neurobiological systems. Some moral decisions have their origins in embodied stress responses associated with increased cortisol, adrenaline, heart rate, and skin conductance (Prasad et al. 2016). Such responses are also frequently associated with affect-processing centers in the brain, such as the limbic system, insula, and amygdala (Moll et al. 2002). By contrast, other decisions seem primarily to originate from or at least be modulated by “cooler” brain regions such as the dorsolateral prefrontal cortex (Decety & Cacioppo 2012). And emerging evidence suggests that the ventromedial prefrontal cortex integrates diverse value inputs from various parts of the brain, transforming them into a common currency or value signal, which guides decision-making (Philiastades et al. 2000, Zaki et al. 2014) that also relate to constructs of the self (Berkman et al. 2017).

Finally, the temporal dimension measures the amount of time that a given type of moral decision-making takes. At one extreme, the most temporally dilated moral psychological research interprets the implications of evolutionary history on a timescale of tens or hundreds of thousands of years, while other research zooms into the scope of ontogenetic development (years or decades), protracted — often social — deliberation (weeks, months, years), and snap decisions (seconds or milliseconds). Moral psychological research by cognitive neuroscientists such as Joshua Greene (2014) and social psychologists such as Jonathan Haidt (2012) has drawn a categorical (not merely discrete or ordinal) distinction between fast, automatic moral mentalizing, on the one hand, and slow, deliberative moral mentalizing on the other hand. Most moral decisions, actions, and reactions are, they argue, driven by the fast system, which is subject to various near-incorrigible biases and irrationalities. Few moral decisions, by contrast, are attributable to the slow system, which can correct some of the errors committed by the fast system but which nevertheless has its own drawbacks and biases. Cushman, Young, & Greene (2010) urge the consolidation of moral psychology around this dual-system framework. And in subsequent work, Cushman (2013 — endorsed by Greene 2017) articulates a more sophisticated dual-process theory in terms of reinforcement learning, which harnesses the distinction between computationally-expensive versus computationally-frugal processes. However these processes are construed, they also tend to map neatly onto the perceived dichotomy between deontology and utilitarianism. The virtue ethical perspective has been largely absent from dual-process moral psychology (see Narvaez, 2015). We will argue that the virtue perspective has much to add to this research program, and that debates focusing entirely on dual process systems overlook important questions about the moral life.

In all of this, we agree with dual-systems proponents that distinguishing the speed of processes is essential, but in this paper we go a step further by drawing comparative distinctions between slower and faster rather than a dichotomous categorical distinction between slow and fast. For example, if two systems dynamically interact over time, the degree to which these systems need to interact, for example when there is a conflict between self-interested and other-oriented motivations, would increase processing time prior to a decision. A virtue theoretic approach would further posit that fluency in moral decisions (more on which below), for example practice in resolving this conflict, could develop it into a habit reflected by speedier decisions (Narvaez, 2010). Dichotomies between fast and slow, model free and model based, utilitarian versus deontological, have a limited view of morality. A virtue theoretic approach argues that moral decisions can be fluent or become more fluent over time and with practice. In addition, multiple perceptual and cognitive capacities are involved in moral actions. These cognitive capacities offer additional points of leverage for moral education.

**3 Two essential virtue ethical constructs: attunement and fluent agency**

In our view, virtue ethical frameworks from both the Western and the Chinese traditions generally share two important constructs, which we will call *attunement* and *fluent agency* (Annas 2011; Slingerland 2014; Stichter 2016). Attunement is a quasi-perceptual capacity related to a valued domain. For example, think of the family of virtues related to generosity, benevolence, altruism, and compassion. Each of these dispositions includes ongoing monitoring and affective responsiveness to (the prospects of) the wellbeing of another person. The generous person is poised to notice opportunities to benefit someone else. The benevolent person is disposed to notice and be gratified by both high absolute levels of and improvements in the levels of wellbeing of others. The altruistic person is liable to notice chances to benefit others, even when it does the altruistic person themselves no good. The compassionate person tracks the wellbeing of others for (the prospects of) decrements and responds negatively to those decrements. To be attuned to the wellbeing of other people in this way is a matter of being disposed to notice some things (e.g., reliable signs of joy or suffering) and not to notice some other things (e.g., the cost to oneself of helping them or mitigating their suffering). It is also a matter of being disposed to experience positive or negative affect, as appropriate, accompanying episodes of such noticing.

Moreover, attunement can be narrower or broader on at least two dimensions: social scale and intensity. You might be attuned to the wellbeing of just a few others, such as your family and close friends. Alternatively, you might be attuned to the wellbeing not just of them but also many others — including even strangers and distant statistical victims. Likewise, you might be highly attuned to the wellbeing of certain others, in the sense that you passively and even actively monitor their wellbeing on a moment-by-moment basis, or less highly attuned, in the sense that you only passively monitor them for strong signals of (changes in) their level of wellbeing but are liable to miss subtler signals. This monitoring need not be conscious or deliberative, and from a virtue theoretic perspective, its automaticity makes it all the more praiseworthy.

 Attunement is thus a standing disposition in an agent that explains a moral agent’s attentional and affective patterns. An agent notices what she does (and doesn’t notice what she doesn’t) *because* she is or isn’t attuned. An agent has an affective response the way she does (and doesn’t respond the way she doesn’t) *because* she is or isn’t attuned. If this is right, then attunement as a disposition precedes voluntary (endogenous) and reactive (exogenous) attention. Further, this attunement precedes affect, and the body and brain structures that harbor attunement must be contiguous with or connected to the structures that embody episodes of attention and affect. Someone could be attuned through some combination of instinct, social pressure or surroundings, acquired habits of attention instantiated in neural architecture through neuroplasticity, and further influences. Regardless of the distal etiology of their attunement, this disposition should be measurable as an individual difference between moral agents, especially in a controlled laboratory environment.

For example, imagine two individuals, A and B, each of whom is to some extent attuned to the wellbeing of a third party, C. If A is more highly attuned than B is, we can make several predictions. First, A should notice signals of C’s wellbeing and prospects *more quickly* than B does. Second, A should notice signals of C’s wellbeing and prospects *more reliably* than B does. Third, in addition to noticing more quickly and more reliably, A should *be rewarded* more by improvements in C’s wellbeing than B is, as well as *be* *punished* more by decrements in C’s wellbeing than B is. Fourth, all else being equal, A should *act* in ways that help or prevent harm to C *more quickly* than B does. Fifth, all else being equal, A should *act* in ways that help or prevent harm to C *more reliably* than B does.

These predictions lead us from attunement to the fluent agency it enables. Fluent agency is an action-guiding capacity related to a valued object, relationship, or domain. Someone exhibits fluent agency to the extent that they find it straightforward to act in ways that promote their values and conflicted when forced or pressured to act in ways counter to their values. Those values could be self-regarding, as in the case of a wholeheartedly greedy person, or other-regarding, as in the case of a wholeheartedly generous person. In the context of the family of virtues currently under discussion (generosity, benevolence, altruism, and compassion), fluent agency is a matter of effectively acting to help and not harm others. Fluent agency in alleviating the suffering of people you can help, on this view, is a matter of being disposed to act in some ways (e.g., to donate resources to someone who could use them) and not to act in other ways (e.g., to refrain from taking resources from those who need them). Someone could exercise fluent agency through some combination of instinct, social pressure and surroundings, acquired habits of action instantiated in neural architecture, and so on. Regardless of the distal etiology, it is difficult if not impossible to exercise fluent agency in some domain without being attuned to that domain. After all, if you don’t even *notice* an opportunity to help, you won’t be in a position to *decide* to help. And if you don’t have a schema for helping, you won’t be in a position to carry out a helping behavior.

Fluent agency thus explains a virtuous agent’s patterns of reward-processing and behavior. An agent experiences the emotions she does (and doesn’t experience the emotions she doesn’t) *because* she embodies attunement toward others. In addition, an agent does what she does (and doesn’t do what she doesn’t) *because* she embodies fluent agency. If this is right, then the temporal and spatial constraints described above mean that attunement precedes reward-processing and action, and that the body and brain structures harboring fluent agency must be contiguous with or connected to the structures that embody episodes of reward-processing and action.

In the case of generosity, for example, someone counts as both attuned and fluent if she is disposed to notice those instances in which she can benefit another person, to act reliably to provide a benefit often enough when she can, not to experience stress when she has to choose between generosity and self-interest, and to find rewarding both benefits to herself and benefits to others. In principle, this network of dispositions could be largely innate, at least for certain beneficiaries, but would be changeable with socialization, education, or intervention. This is more or less Mengzi’s conception of the sprout of *ren* (Slingerland 2014; Wong 2013), which disposes its bearer to be compassionately attuned to the suffering of close friends and family, as well as cute animals, but not all conspecifics. For Mengzi, the sprout of *ren* develops into a virtue via patient processes of social and reflexive enculturation.[[3]](#footnote-4)

Because the family of virtues under consideration are “low-fidelity” virtues that can be fully embodied even if they do not eventuate in action in every relevant circumstance (Alfano 2013) -- in other words there can be variability in responses even in a virtuous person -- it would be inappropriate to operationalize them in terms of strict behavioral consistency. However, it does seem apt to say that a generous person will help more quickly and more reliably than a less generous person. In addition, we predict that a more generous person should regulate their own motivations and emotions in the face of value conflicts more effectively than someone who is not wholeheartedly generous. So, if the reflections on attunement and fluent agency are on the right track, they pave the way for integrating a virtue ethical perspective into accounting for individual differences, while avoiding overly simplistic dichotomies. More specifically, the constructs of attunement and fluent agency point to multiple comparative predictions about the speed and reliability of attentional and decisional processes, about the magnitude of reward processing, and about the regulation of affective processes. We have taken the family of virtues encompassing generosity, benevolence, altruism, and compassion as our case study, but presumably similar predictions could be made about other virtues such as honesty, curiosity, courage, and fairness. In this way, the temporal domain becomes a measure of virtue when applied to moral motives, decisions and behavior. Importantly, this is quite a different approach than the way the temporal ordering is used in moral theories that rely on a dichotomous perspective, which we outline in the next section.

**4 Limitations of the fast/slow and hot/cold dichotomies**

 To review, a virtue theoretic approach would say that moral actions that are more reflexive, that require less deliberation or conflict can be praiseworthy. A person who intuitively and reflexively acts with benevolence has codified moral actions into habits, and from an Aristotelian perspective has developed fluency, not mere continence. Typical dual-process theories map deontological rules onto overlearned automatic affective processes and utilitarianism onto slower, deliberative processes (with a recent effort by some to relate these rules to virtue (Greene 2017)). Here, we review the empirical literature relevant to the constraints and predictions catalogued above, and explain why we are advocating for an alternative to the fast/slow and hot/cold dichotomies in contemporary psychological science (Kahneman 2011), including some contemporary moral psychology (Haidt 2012; Greene 2013, 2014). The basic idea is that these dichotomies are often not sufficiently fine-grained to test the predictions about temporal and spatial order, particularly when examining the questions of attunement and fluency.

As we previously described, dual-process theories tend to distinguish categorically between fast, affect-laden, automatic, sometimes-unconscious, nearly-incorrigible mental processes and slow, rational, effortful, conscious, corrigible mental processes, mapping deontological considerations onto the fast and consequentialist considerations onto the slow. It is important to note however, that the temporal scale that distinguishes the “fast” from the “slow” is rarely if ever made explicit. Presumably, the distinction is comparative: relative to deliberative processing, emotional responses or intuitions are faster. But how fast, in seconds or milliseconds, does a process have to be to count as “fast”? And if a neural structure with a relatively fast response — say the response of the amygdala —modulates a second process, should we still label this second process as “fast”?

In addition, certain methodological constraints in assigning fast or slow to these processes are not always acknowledged. Researchers often seem to assume that the temporal resolution of reading paragraph-length vignettes while being scanned by functional magnetic resonance imaging (fMRI) is sufficiently fine-grained to distinguish fast from slow. In particular, Greene (2013, pp. 119-27, 295-6) has argued that moral dilemmas like the trolley problem (Foot 1967; Thomson 1976; Kamm 1989; Unger 1996) are moral variants of the Stroop task (Stroop 1935), one of the oldest and best-established tasks in scientific psychology. Below, we argue that this analogy is not apt.

The trolley problem was originally posed as a hypothetical dilemma to illustrate the moral difference between doing harm and allowing harm. Foot asks her readers to imagine they find themselves at the operator’s switch when they see a runaway trolley headed towards five workers who will not be able to escape the tracks before the trolley overtakes and kills them. On a spur track is a single worker. You could pull the switch and divert the trolley onto the spur track, saving the five workers but leading to the death of the one. Most people encountering this hypothetical dilemma for the first time agree with Foot that the right thing to do is pull the switch and divert the trolley (Cushman et al. 2008). Foot then invites her readers to consider a variant of the dilemma, in which the protagonist can save the five workers by pushing a very large man off a footbridge into the path of the trolley. This man will be killed, but his bulk will be sufficient to stop the trolley and save the five. Most people encountering this hypothetical dilemma for the first time agree with Foot that the right thing to do is refrain from pushing the large man, thereby allowing the five workers to die. The difficulty, of course, is that — on their face — both dilemmas involve a tradeoff between five lives and one life. The gruesomeness of the footbridge scenario shouldn’t, one might think, make a moral difference. According to adherents of the dual-process framework, the difference in people’s intuitions is explained by the fact that, when considering the *switch* dilemma, people engage in slower, affect-free comparisons, whereas when they consider the *footbridge* dilemma, they employ fast processes laden with negative affect. These alleged fast, affective responses short-circuit their capacity to engage in reflective comparisons and produce the psychologically-understandable but morally-indefensible intuition that it is wrong to push the large man.

This kind of interference, in which one process makes it more difficult to effectively rely on another process, is the mechanism that explains the Stroop effect. The basic finding with the original Stroop task is that people are able to read color-words printed in congruently-colored ink faster than they are able to read color-words printed in incongruently-colored ink. For example, if you are literate in English, you will tend to name the color of the ink in the word ‘**black**’ (the ink is black) more quickly than the color of the ink in the word ‘**white**’ (the ink is black). The automatic and overlearned semantic process of reading a single printed word interferes with naming the color of the ink. As MacLeod (1991) shows in a review of five decades of research, this effect is robust and has been replicated many times, with reported reaction times for congruent and incongruent trials typically between 200 milliseconds and 800 milliseconds.

There are several variants of the Stroop task, such as one that pits the recognition of emotional facial expressions against words that refer to emotions (Stenberg, Wiking, & Tahl 1998). People are, for example, able to read the word ‘angry’ superimposed over an angry face more quickly than they are able to read the word ‘angry’ superimposed over a sad face. What unifies the variants of the Stroop task is their mechanism: they pit mental processes that take comparable amounts of time against each other under speeded conditions, and when they deliver incongruent results to a shared process, they are forced to compete and a winner is named. The neural response to this interference peaks at about a half second after the stimulus (Liotti, Woldorff, Perez, & Mayberg, 2000; West & Alain, 2000). This timing is comparable to objectively “fast” processes; for example, the blink of an eye takes about 300 milliseconds and a batter must react reflexively to the 400 millisecond journey of a fastball from a baseball pitcher’s hand to home plate. However, perceptual processing and its modulation by attention can occur even faster than 100 ms (e.g. Karns & Knight 2009).

One process (P1) can only interfere with another process (P2) if it occurs before P2 or simultaneously with P2. The Stroop task and its variants are therefore a way of establishing temporal priority or simultaneity, and therefore a way of respecting the constraints established in the previous section. In theoretical models and simulations of such interference, it’s typically assumed that P1 and P2 are initially processed in parallel by different (though potentially overlapping) populations of neurons before converging on the process that registers interference (MacLeod 1991) — potentially in the anterior cingulate cortex (Liotti et al. 2000). If P1 occurs before P2, it should be possible to show that P1 interferes with P2 but not vice versa. Likewise, if P2 occurs before P1, it should be possible to show that P2 interferes with P1 but not vice versa. And if they occur simultaneously, it should be possible to show that they both interfere with each other. In the case of colored ink and word meaning, there is a large body of evidence for this interference. However, Stroop-type tasks with moral relevance are rare. We are aware of two published papers in which the color of text (white/black — assumed to have good/evil moral valence) was pitted against the moral valence of the referents of the words to discover a Stroop-type effect for moral processing (Sherman & Clore 2009; Chiou & Cheng 2013).[[4]](#footnote-5) In keeping with more traditional Stroop tasks, both of these studies report reaction times for individual trials of between 200 milliseconds and 800 milliseconds, which suggests that they are measuring processes that are (nearly) matched for speed.

Let us now return to the arguments advanced by Greene, Haidt, and others advocating a dual-process framework for moral judgment and decision-making. The stimuli used in these studies *almost never* elicit processes embodying the (near-)simultaneity needed to produce a Stroop-like effect. In Haidt’s studies, participants read a sentence or paragraph designed to elicit an emotion such as contempt, anger, or disgust. Then they read a question asking them to express a moral judgment. For example, they read a fictional story of sibling incest in which the siblings make significant efforts to hide their one-off encounter from others, use effective birth control, enjoy the experience, and never regret what they did. In Greene’s studies, participants read a paragraph-length description of the *switch* or *footbridge* version of the trolley dilemma. They may have the opportunity to cross-reference the prose with a figure representing the story. Participants can ponder how to respond for as long as they like or, in studies that involve time pressure, several seconds (Greene et al. 2004). Then they choose a response that expresses a moral judgment.

All told, even a relatively quick response to the sort of stimulus typically used to study “fast” moral cognition takes 10 to 100 times as long as a response to a standard Stroop task. During the time it takes to read and respond to a vignette, a participant is liable to experience dozens of both intuitive and reflective evaluations and re-evaluations of (parts of) the stimulus material.[[5]](#footnote-6) There is just no way that this kind of stimulus pits one process of one type against one other process of another type in some sort of race. Instead, given what we know about typical human reaction times, it pits a host of processes of various types against another host of partially-overlapping processes of various types. Perhaps the more conflicting a scenario, the more times it must be simulated in the mind’s eye to arrive at a decision. A Stroop-like pairwise interference or non-interference is impossible to establish under these conditions.

To make matters worse, the methodology typically used to measure the temporal dimension of brain activation in trolley cases (fMRI) lacks the temporal resolution needed to directly answer questions about fast versus slow comparisons. Stroop-type interference occurs 200-800 milliseconds after stimulus onset. The temporal resolution of fMRI is seconds or even minutes (Sejnowski et al. 2014). As Bryce Huebner (2015, p. 3) puts it

studies examining *moral judgments* move at a glacial pace relative to the speed of neural processing. Reading a moral dilemma, and coming to a decision about its permissibility, may take a long time. The time scale of these experiments should trigger numerous cognitive processes, which will interact at various time-scales.

For example, the increased blood-oxygen-level dependent (BOLD) signal in “emotion” areas reported in Greene (2001) could indicate a greater aggregated amount of processing in these regions over an extended period of time, as the simulation of the dilemma is run in the participant’s mind, rather than a fast or automatic intuition. Although there is evidence from various domains that the amygdala does respond quickly to brief events, as early as 74 ms for fearful faces (Méndez-Bértolo et al. 2016), there is little evidence that, in the case of protracted moral reasoning, it plays a similar reactive role. This suggests that genuinely distinguishing the speed of moral cognition requires either improvements in fMRI’s temporal resolution or alternative methodologies, such as electrophysiological measures that have faster temporal resolution, such as EEG.

If the criticisms offered thus far are on the right track, we can draw some preliminary conclusions. First, insofar as fast processes are allied with deontological intuitions and slow processes with utilitarian intuitions, virtue ethical constructs and individual differences in fluent virtues are often excluded from dual-process moral psychology. Second, many prominent studies of allegedly fast moral mentalizing - let’s assume “fast” would be Stroop-like here - use stimuli that are only appropriate for studying slow moral mentalizing. Third, many prominent studies of allegedly fast moral mentalizing use methods that could not, even in principle, detect fast mentalizing. Fourth, many prominent studies of allegedly fast moral mentalizing do not measure the interaction of one fast process with another fast process; instead, they measure the aggregate outputs and interactions of dozens of processes — some of which might be characterized as fast, others of which might be characterized as slow. Fifth, advocates of the fast/slow dichotomy in moral psychology have thus far failed to suggest a temporal threshold to distinguish fast from slow moral mentalizing, making the mapping of deontology to fast and consequentialism to slow problematic.

Perhaps this is at the crux of the problem, time is continuous, and fast versus slow distinctions are categorical. Let us call this the problem of Underspecified Temporal Quantities (UTQ): what counts as fast or slow is not specified in advance, making it difficult to tell whether any particular measure tracks a fast or a slow process. We propose that these fast/slow distinctions should be applied more cautiously. With a more fine-grained temporal process we would be able to make more precise claims about which processes are faster than others. For example, a social psychologist interested in quick flashes of affect that individuals experience in the face of emotionally and morally arousing stimuli may be measuring processes that take between 100 milliseconds and 2 seconds, and given a specific experimental design, are testing their temporal hypothesis at a specific social and spatial scale.

When these social, spatial, and temporal precisifications have been made, the scientist is in a position to decide what methodology is appropriate to measure the constructs under study. The scientist should only employ a methodology that can effectively discriminate *within* the specified social, spatial, and temporal intervals. You wouldn’t use a yardstick to compare the diameters of people’s pupils: it couldn’t distinguish them effectively. You wouldn’t use a truck weighing station to compare the weights of two human adults: it couldn’t distinguish them effectively. For the same reason, we contend, you shouldn’t use fMRI to compare the priority of mental processes in a single individual’s brain: it can’t distinguish them effectively, though it is excellent for spatial measurements. Instead, moral psychologists interested in studying “fast” processes should use a methodology that can distinguish at least at the level of several milliseconds. The most prominent such methodologies that can be used on human subjects are electroencephalography (EEG) and magnetoencephalography (MEG).

A review of the literature indicates that the few papers that have employed EEG used materials in which participants read or saw intentional or accidental harms or benefits to someone else.[[6]](#footnote-7) For instance, Decety & Cacioppo (2012) found that intentionality is processed in the brain (right pSTS, 60-70 ms) before negative affect (amygdala/temporal pole, 120-130 ms), which is processed before value-integration (vmPFC, 180-90 ms). Yoder & Decety (2014) found that moral valence is encoded before 300 ms, and Xue, Wang, & Tang (2013) and Sarlo et al. (2014) used EEG to monitor participants reading and forming moral judgments about morally dilemmatic vignettes. Arguably EEG studies in neuroeconomics such as Declerck & Boone (2016; see also Strobel 2016) should be counted among the class of research exemplifying the methodological constraints we are recommending, as should EEG research on charitable giving that does not explicitly label itself “moral psychology” but is relevant to generosity, such as San Martín et al. (2016). In summary, researchers in the field of moral psychology have mostly failed to be clear how small and how fast the processes that they want to study are, which has led them to employ methodologies that often cannot discriminate effectively within the spatial and temporal parameters of the phenomena of interest.

 Attunement and fluency – two concepts central to virtue theory – could be measured in the context of the more fine-grained framework we have been advocating. Such an approach helps to move beyond overly simplistic dichotomies between fast/slow or deontological/utilitarian. In Section 3, we claimed that attunement is characterized as a disposition which explains an agent’s attentional and affective patterns. This was further elaborated in terms of two claims: the agent notices what she does *because* she is attuned to some features of her environment (and not others); the agent feels what she does *because* she is attuned to some features of the environment (and not others). How, then, might we measure attunement so-construed?

 Recalling Figure 1, there are three dimensions along which careful distinctions should be made: temporal, spatial, and social. To see the importance of making such distinctions, consider the temporal dimension. Researchers might be interested in trait-like attunement, in which case measurements of slower, more reflective processes (such as those elicited by questionnaires or resting state methods) would be appropriate. But researchers might also be interested in particular time slices, or state-like manifestations of attunement, in which case measurements of faster, more automatic process (such as those elicited by speeded tasks, or tasks performed under cognitive load) would be appropriate. Alternatively, researchers may be interested in how various state-like manifestations of attunement coalesce into more trait-like dispositions over the lifespan, in which case longitudinal designs would be appropriate. And given that the temporal quantities described here range from the order of milliseconds to years, it is crucial to specify the relevant timescale in advance and provide reference points and examples of fast versus slow processes in the context where these processes are being discussed.

**5. Conclusions**

 To sum up, we began by laying out a multidimensional framework for moral psychological inquiry, with a particular focus on the temporal dimension. In Section 3, we then spelled out the virtue ethical constructs of attunement and fluent agency, arguing that these constructs commit the virtue ethicist to several empirically tractable because-claims that entail temporal, spatial, and social constraints. In Section 4, we then showed how these constraints are often difficult to implement in research employing a dichotomous distinction between for example, slow and fast mental processes. We concluded that these kinds of temporal claims would benefit from increased precision and pre-specification.

There are three important points that follow from our arguments here. First, we hope to have shown how a virtue-theoretic approach offers a plausible and attractive way to extend our thinking about moral psychology beyond dichotomies. This is not a sweeping condemnation of all dual-process models, of course, but rather a suggestion that a rich and robust moral psychology should be able to account for the cross-cutting, and overlapping nature of various cognitive and affective processes, the biological mechanisms underpinning them, and the varied timescales on which they occur.

Second, by focusing on some of the methodological shortcomings of dichotomous thinking, particularly along the temporal dimension, we have also tried to show how a virtue-theoretic approach is well-suited to methodologically sound and theoretically robust moral psychology.

Third, by calling attention to the virtue ethical constructs of attunement and fluent agency, we hope to have shown that the scope of moral psychological inquiry is broader than the literature on moral dilemmas might suggest. The ways in which moral agents can be attuned (or not) to various features of their environment highlights the importance of developmental cognitive, affective and motivational capacities. In turn, this expands the possibilities and varieties of moral education and moral cultivation.

**6. References**

Alfano, M. (2013). *Character as Moral Fiction*. Cambridge University Press.

Alfano, M. (2016a). *Moral Psychology: An Introduction*. Polity.

Alfano, M. (2016b). Friendship and the structure of trust. In A. Masala & J. Webber (eds.), *From Personality to Virtue: Essays in the Psychology and Ethics of Character*, 186-206. Oxford University Press.

Alfano, M. (2017). The topology of communities of trust. *Russian Sociological Review*, 15(4): 30-56.

Alfano, M., Huijts, N., & Roeser, S. (forthcoming). Trust and distrust in institutions and governance. In J. Simon (ed.), *Handbook of Trust and Philosophy*. Routledge.

Alfano, M. & Skorburg, A. (2017). The embedded and extended character hypotheses. In J. Kiverstein (ed.), *Handbook of Philosophy of the Social Mind*. Routledge.

Annas, J. (2011). *Intelligent Virtue*. Oxford University Press.

Berkman, E. T., Livingston, J. L., & Kahn, L. E. (2017). Finding the “self” in self-regulation: The identity-value model. *Psychological Inquiry*, *28*(2-3), 77-98.

Besser-Jones, L. (2014). *Eudaimonic Ethics: The Philosophy and Psychology of Living Well*. Routledge.

Christen, M. & Alfano, M. (2013). A research program for empirically informed ethics. In M. Christen et al. (eds.), *Empirically Informed Ethics: Morality between Facts and Norms*, 3-27. Springer.

Chiou, W.-B. & Cheng, Y.-Y. (2013). In broad daylight, we trust in God! Brightness, the salience of morality, and ethical behavior. *Journal of Environmental Psychology*, 36: 37-42.

Cushman, F., Knobe, J., & Sinnott-Armstrong, W. (2008). Moral appraisals affect doing/allowing judgments. *Cognition*, 108(2): 353-80.

Cushman, F., Young, L., & Greene, J. (2010). Our multi-system moral psychology: Towards a consensus view. In J. Doris et al. (eds.), *The Oxford Handbook of Moral Psychology*. Oxford University Press.

Decety, J. & Cacioppo, S. (2012). The speed of morality. *Journal of Neurophysiology*, 108(11): 3068-72.

Declerck, C. & Boone, C. (2016). *Neuroeconomics of Prosocial Behavior: The Compassionate Egoist*. Elsevier.

Dunbar, R. (1992). Neocortex size as a constraint on group size in primates. *Journal of Human Evolution*, 22(6): 469-93.

Dunbar, R. (1993). Coevolution of neocortex size, group size, and language in humans. *Behavioral and Brain Sciences*, 16(4): 681-735.

Foot, F. (1967). *Virtues and Vices*. Blackwell.

Francis, K., Howard, C., Howard, I., Gummerum, M., Ganis, G., Anderson, G., & Terbeck, S. (2016). Virtual morality: Transitioning from moral judgment to moral action? *PLoS ONE* 11(10): e0164374. doi:10.1371/journal.pone.0164374.

Gray, K., Waytz, A., & Young, L. (2012). The moral dyad: A fundamental template unifying moral judgment. *Psychological Inquiry*, 23(2): 206-15.

Greene, J. (2008). The secret joke of Kant’s soul. In W. Sinnott-Armstrong (ed.), *Moral Psychology vol. 3: The Neuroscience of Morality: Emotion, Brain Disorders, and Development*. MIT Press.

Greene, J. (2013). *Moral Tribes: Emotion, Reason, and the Gap between Us and Them.* Penguin.

Greene, J. (2014). Beyond point-and-shoot morality: Why cognitive (neuro)science matters for ethics. *Ethics*, 124(4): 695-726.

Greene, J. (2017). The rat-a-gorical imperative: Moral intuition and the limits of affective learning. *Cognition*.

Greene, J., Nystrom, L., Engell, A., Darley, J., & Cohen, J. (2004). The neural bases of cognitive conflict and control in moral judgment. *Neuron*, 44(2): 389-400.

Haidt, J. (2012). *The Righteous Mind: Why Good People are Divided by Politics and Religion*. Random House.

Holyoak, K. & Powell, D. (2016). Deontological coherence: A framework for commonsense moral reasoning. *Psychological Bulletin*, 142(11): 1179-1203.

Karns, C. M., & Knight, R. T. (2009). Intermodal auditory, visual, and tactile attention modulates early stages of neural processing. *Journal of Cognitive Neuroscience*, *21*(4), 669-683.

Kamm, F. (1989). Harming some to save others. *Philosophical Studies*, 57: 227-60.

Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3): 726-36.

Lakoff, G. & Johnson, M. (1999). *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. Basic Books.

Liao, M. (2016). *Moral Brains: The Neuroscience of Morality*. Oxford University Press.

Liotti, M., Woldorff, M. G., Perez, R., & Mayberg, H. S. (2000). An ERP study of the temporal course of the Stroop color-word interference effect. *Neuropsychologia*, 38: 701–711.

Lurquin, J. & Miyake, A. (2017). Challenges to ego-depletion research go beyond the replication crisis: A need for tackling the conceptual crisis. *Frontiers in Psychology*, 8(568): 1-5.

MacLeod, C. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2): 163-203.

Martín, R., Kwak, Y., Pearson, J., Woldorff, M., & Huettel, S. (2016). *Social Cognitive and Affective Neuroscience* 11(6): 863-76.

Méndez-Bértolo, C., Moratti, S., Toledano, R., Lopez-Sosa, F., Martínez-Alvarez, R., Mah, Y.H., Vuilleumier, P., Gil-Nagel, A. and Strange, B.A., (2016). A fast pathway for fear in human amygdala. *Nature neuroscience*, *19*(8), 1041-1049.

Moll, J., de Oliveira-Souza, R, Eslinger, P., Bramati, I., Mourão-Miranda, J., Andreiuolo, P. A., & Pessoa, L. (2002). Functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience*, 22(7): 2730-36.

Narvaez, D. (2010). Moral complexity: The fatal attraction of truthiness and the importance of mature moral functioning. *Perspectives on Psychological Science, 5*(2), 163-181*.*

Narvaez, D. (2015). The co-construction of virtue: Epigenetics, neurobiology and development. In N. E. Snow (Ed.), *Cultivating Virtue* (pp. 251-277). New York, NY: Oxford University Press.

Narvaez, D., & Lapsley, D.K. (2005). The psychological foundations of everyday morality and moral expertise. In D.K. Lapsley & Power, C. (Eds.), *Character Psychology and Character Education* (pp. 140-165). Notre Dame: IN: University of Notre Dame Press.

Philiastides, M., Biele, G., & Heekeren, H. (2010). A mechanistic account of value computation in the human brain. *Proceedings of the National Academy of Sciences*, 107(20): 9430.

Prasad, S., Narayanan, J., Lim, V., Koh, G., Koh, D., & Mehta, P. (2016). Preliminary evidence that acute stress moderates basal testosterone’s association with retaliatory behavior. *Hormones and Behavior*, 92: 128-40.

Railton, P. (2014). The affective dog and its rational tale: Intuition and attunement. *Ethics*, 124(4): 813-59.

Sarlo, M., Lotto, L., Rumiati, R., & Palomba, D. (2014). If it makes you feel bad, don’t do it! Egoistic rather than altruistic empathy modulates neural and behavioral responses in moral dilemmas. *Physiology & Behavior*, 130: 127-34.

Sherman, G. & Clore, G. (2009). The color of sin: White and black are perceptual symbols of moral purity and pollution. *Psychological Science*, 20(8): 1019-25.

Singer, P. (1972). Famine, affluence, and morality. *Philosophy and Public Affairs*, 1(3): 229-43.

Slingerland, E. (2014). *Trying Not to Try: The Art and Science of Spontaneity*. Crown.

Snow, N. (2006). Habitual virtuous actions and automaticity. *Ethical Theory and Moral Practice*, 9(5): 545-61.

Snow, N. (2009). *Virtue as Social Intelligence: An Empirically Grounded Theory*. Routledge

Stenberg, G., Wiking, S., & Dahl, M. (1998). Judging words at face value: Interference in a word processing task reveals automatic processing of affective facial expressions. *Cognition and Emotion*, 12(6): 755-82.

Stichter, M. (2016). Practical skills and practical wisdom in virtue. *Australasian Journal of Philosophy*, 94(3): 435-48.

Strobel, A. (2016). *Neuroeconomics*. Springer.

Strohminger, N., Caldwell, B., Cameron, D., Schaich Borg, J., & Sinnott-Armstrong, W. (2014). Implicit morality: A methodological survey. In C. Luetge, H. Rusch, & M. Uhl (eds.), *Experimental Ethics: Towards an Empirical Moral Psychology*. Palgrave Macmillan.

Stroop, J. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6): 643-62.

Thomson, J. J. (1976). *Killing, Letting Die, and the Trolley Problem*, 59: 204-17.

Unger, P (1996). *Living High and Letting Die*. Oxford University Press.

Västfjall, D., Slovic, P., Mayorga, M., & Peters, E. (2014). Compassion fade: Affect and charity are greatest for a single child in need. *PLoS ONE*, 9(6).

West, R., & Alain, C. (1999). Event-related neural activity associated with the Stroop task. *Cognitive Brain Research*, 8: 157–164.

Wong, D. (2013). Chinese ethics. *The Stanford Encyclopedia of Philosophy* (Spring 2013 edition), E. Zalta (ed.), URL = <http://plato.stanford.edu/archives/spr2013/entries/ethics-chinese/>.

Xue, S.-W., Wang, Y., & Tang, Y.-Y. (2013). Personal and impersonal stimuli differentially engage brain networks during moral reasoning. *Brain and Cognition*, 81(1): 24-28.

Yoder, K. & Decety, J. (2014). Spatiotemporal neural dynamics of moral judgment: A high-density ERP study. *Neuropsychologia*, 60: 39-45.

Zaki, J., López, G., & Mitchell, J. (2014). Activity in ventromedial prefrontal cortex co-varies with revealed social preferences: Evidence for person-invariant value. *SCAN*, 9(4): 464-9.

1. See Holyoak & Powell (2016) for more. [↑](#footnote-ref-2)
2. The Dunbar number refers to the predicted maximal group size of small-band hunter-gatherers, as predicted by human neocortex volume. While it is clear that humans can form stable groups larger than this, they only did so relatively recently in evolutionary terms, and many contemporary groups, such as churches and effective military units, continue to face an upper bound of approximately 150 members. [↑](#footnote-ref-3)
3. For more on this line of thought, see Snow (2006 & 2009). [↑](#footnote-ref-4)
4. Moral valence has also been pitted against direction, based on purported metaphorical equivalence of *good* with *up* and of *bad*/*evil* with *down*. However, Lakens (2012) has persuasively argued that there is a simpler, non-moral explanation for this effect based simply on polarity correspondence. [↑](#footnote-ref-5)
5. For more on this point, see Railton (2014), Narvaez (2010), and Narvaez & Lapsley (2005). [↑](#footnote-ref-6)
6. This literature review was conducted on Google Scholar for the conjunction of ‘EEG’ and ‘moral’— search performed 27 May 2017. [↑](#footnote-ref-7)