Folk Theories in the Moral Domain

Sara Gottlieb

Tania Lombrozo

University of California, Berkeley

Abstract

Is morality intuitive or deliberative? This distinction can obscure the role of folk moral theories in moral judgment; judgments may arise "intuitively" yet result from abstract theoretical and philosophical commitments that participate in "deliberative" reasoning.

Introduction and historical context

Physician-assisted suicide was, according to the Gallup Poll, the most controversial social issue in 2010: 46% of the individuals surveyed indicated that it was morally acceptable, and a matching 46% indicated that it was morally wrong ("Four Moral Issues Sharply Divide Americans," 2010). It remained controversial in 2012 when Massachusetts voted against the Death With Dignity Act by a narrow 1%. And in 2014, the issue regained momentum in the media with Brittany Maynard, the terminally ill 29 year-old who publicly documented the decision to take her own life.

Individuals support or oppose physician-assisted suicide for a variety of reasons. The American College of Physicians put forth an official stance in a 2001 position paper, stating that although "arguments supporting physician-assisted suicide highlight the duty to relieve patient suffering or stem from a vigorous understanding of the duty to respect patient autonomy" (p. 211), the Hippocratic Oath requires that physicians follow a tradition of healing and comfort and never intentionally bring about the death of any patient. They wrote, "Just as society can direct that no one has the 'right' to sell himself or herself into slavery, so too can society direct that no one has a 'right' to assistance with suicide" (p. 212).

This example illustrates one process by which moral judgments can be reached: through the explicit consideration and weighing of relevant moral principles, such as respecting patient rights, or adhering to natural law concerning time of death. But is this how the respondents to the 2010 Gallup Poll reached their judgments, as well? Traditional and more contemporary accounts of moral judgment offer different responses. According to more traditional accounts, such as those grounded in classic work by Kohlberg (Kohlberg, 1996; Turiel, 1983), moral judgments typically result from a process of explicit moral reasoning – or "deliberation" – akin to that

offered by The American College of Physicians. More recent accounts, such as Haidt's (2001) social intuitionist model, however, challenge the idea that moral justifications are causally responsible for their corresponding judgments. Instead, they argue that moral judgment is a fundamentally "intuitive" phenomenon, with a large literature suggesting that moral attitudes on issues related to sanctity of life, which typically divide liberal and conservative voters, are guided by *affect* – most notably disgust (Inbar, Pizarro, & Bloom, 2009; Inbar, Pizarro, Knobe, & Bloom, 2009; Inbar, Pizarro, Iyer, & Haidt, 2012; Inbar, Pizarro, & Bloom, 2012).

These opposing approaches differ critically on the emphasis they place on *deliberative* versus *intuitive* processes, a distinction that has paved the way for widespread "dual-systems" or "dual-process" accounts of moral reasoning (Greene, 2007; Greene, Sommerville, Nystrom, Darley, & Cohen, 2001; Greene, Morelli, Lowenberg, Nystrem, & Cohen, 2008; Greene, Nystrom, Engell, Darley, & Cohen, 2004). In this chapter we argue that while dual-systems approaches have been useful in many ways, a sharp boundary between intuition and deliberation potentially obscures important phenomena in moral judgment. In particular, we argue that in many cases, moral judgments can arise "intuitively" yet result from abstract and coherent theoretical commitments that participate in "deliberative" reasoning. One example comes from the case of physician-assisted suicide, which implicitly ties a terminally ill patient's deciding mind to her failing body. For this and other issues that bear on the sanctity of life, moral judgments could depend not only on affect, but on relatively abstract and coherent metaphysical commitments concerning the relationship between the mind and the body – what has typically been referred to as "intuitive dualism" in the psychological literature (Bloom, 2004; Greene, 2006). We propose that theoretical commitments like those embodied in intuitive dualism play an important role in moral judgment, but in a manner that crosscuts the traditional

intuition/deliberation divide. This hybrid proposal borrows insights from both traditional and contemporary accounts of moral judgment from social psychology, but also draws on research from both cognitive and developmental psychology on intuitive theories of the natural world.

To argue for this proposal, we first provide a brief review of evidence that has been used to support the dual-process perspective in moral psychology. We then suggest that the distinction between intuitive and deliberative processing is potentially problematic when it comes to describing the role of more abstract commitments in moral judgment, as in the example of physician-assisted suicide and intuitive dualism. To make sense of such cases we turn to literature on intuitive theories in other domains and argue for "folk theories" that play a role in shaping moral judgment. We then present evidence for this position, including our own recent work, which documents systematic relationships between people's metaphysical and epistemic commitments, on the one hand, and their intuitive judgments concerning bioethical issues such as physician-assisted suicide, on the other.

Intuition versus deliberation: A dual-process perspective

Dual process theories, of which there are many (Evans & Stanovich, 2013), typically differentiate two types of thinking: one intuitive and the other deliberative. Intuition and deliberation typically map onto mental processes underlying decision-making and behavior differentiated according to whether they operate automatically or in a controlled manner. This distinction, in turn, can be operationalized either behaviorally – with automatic processes manifesting themselves under cognitive load or time pressure – or by isolating distinct neural correlates (e.g., the VMPFC versus DLPC for automatic versus controlled processes in the case of deontological versus utilitarian judgment, Greene et al., 2001; 2004). The distinction between

automatic and controlled processes, and correspondingly between intuitive and deliberative judgments, has been particularly influential in moral psychology.

While there are many flavors of dual-process theories (see Evans & Stanovich, 2013), most agree in linking intuitive versus deliberative processing with the pairs of opposing attributes identified in Table 1.

	Intuitive	**Deliberative**
Process	Automatic	Controlled
Speed of processing	Fast	Slow
Role of affect	Often high	Often low
Level of consciousness	Nonconscious	Conscious
Representation	Contextualized	Abstract
Accuracy	"Good enough"	Often high
Evolutionary origin	Distant	Recent
Type of belief	Implicit	Explicit

Evidence for the distinction between intuitive and deliberative processing accordingly focuses on these attributes, with particular emphasis on the first three.

Initial support for dual-process approaches to the moral domain came from functional neuroimaging studies investigating the extent to which brain processes associated with emotion (e.g., VMPFC, amygdala) are engaged in response to different kinds of moral dilemmas (Greene et al., 2001; 2004). For example, Greene and colleagues presented participants with variants on trolley car problems, such as the hypothetical footbridge case, in which a participant must decide whether it is permissible to push one person in front of a train to prevent the train from hitting five others. Scenarios of this sort create a tension between deontological bases for judgments, which reflect rights and duties, and utilitarian bases for judgment, which require favoring the

greater good. Greene et al. (2001) found that "personal" moral dilemmas – those like the footbridge case that involve causing direct harm, often through touch – tended to elicit neural activity associated with emotion. In contrast, Greene et al. (2004) found evidence for brain processes associated with cognitive control (e.g., DLPFC) in utilitarian moral judgment. These initial findings supported the idea that deontological judgments emerge from "intuition" (with an important role for automatic emotional processing), and utilitarian judgments from more controlled deliberation.

Subsequent work has backed up the association between deontology and more intuitive processing, on the one hand, and between utilitarian judgments and deliberation, on the other (Paxton, Ungar, & Greene, 2011). For example, patients with fronto-temporal dementia (characterized by "emotional blunting") are three times more likely than healthy controls to answer in favor of pushing the man off the footbridge for utilitarian benefit (Mendez, Anderson, & Shapira, 2005). At the cellular level, citalopram, a selective serotonin reuptake inhibitor (SSRI), increases the availability of serotonin in the blood stream, thereby increasing certain emotional responses and deontological moral judgment, while anti-anxiety drugs like lorazepam can reduce deontological inclinations (Perkins, Leonard, Weaver, Dalton, Mehta, Kumari, Williams, & Ettinger, 2013). And at a behavioral level, utilitarian judgments are more affected by cognitive load (Greene et al., 2008; Tremoliere, Neys, & Bonnefon, 2012), associated with longer decision time (Suter & Hertwig, 2011), and associated with reflective, as opposed to intuitive, mindsets (Paxton et al., 2011).

One implication of dual-process approaches is that moral judgments (which reflect a mix of intuition and deliberation) can seriously depart from moral justifications (which fall on the side of deliberation). And in fact, there's evidence that the two diverge. Haidt (2001), for

example, argues for a phenomenon he calls "moral dumbfounding," which refers to an individual's inability to produce moral justifications for moral judgments. As evidence, he presents the case of Mark and Julie – siblings who decide to engage in consensual sex, use protection, and find that it brings them closer together. An overwhelming number of individuals find this wrong, but when probed why, fail to produce reliable justifications. Haidt (2001) claims that "moral reasoning does not cause moral judgment; rather, moral reasoning is usually a post hoc construction, generated after a judgment has been reached" (p. 814). Other empirical work similarly suggests that moral justifications are not responsible for their corresponding moral judgments (Hauser, Cushman, Young, Jin, & Mikhail, 2007), and that even when individuals can provide justifications, they sometimes fail to recognize the full set of factors that influenced their judgments (Cushman, Young, & Hauser, 2006).

Dual-systems approaches can also accommodate cases in which moral judgments and moral justifications systematically cohere – the approach does not reject the possibility that moral justifications can sometimes influence judgments, or that justifications will match judgments because they are generated post-hoc. A bigger challenge for most dual-systems approaches would come from evidence for representational structures that blur the crucial distinction between "intuitive" and "deliberate" processing. A candidate for such a structure comes from research in cognitive and developmental psychology that aims to characterize people's intuitive theories of the natural world, such as folk psychological, folk biological, and folk physical beliefs. As we detail in the next section, "intuitive" theories don't fit neatly on a single side of the intuitive/deliberative divide.

Drawing an analogy to folk scientific theories

A broad literature in cognitive and developmental psychology suggests that children hold rich intuitive theories of the world even before they begin formal education (e.g., Carey, 2000; Keil, 2011). In the domain of physics, for example, students hold theories grounded in the belief that forces transfer from one object to another upon contact and must dissipate before those objects cease moving (Clement, 1982; McCloskey, 1983). In the domain of biology, children hold intuitive theories of adaptation grounded in a belief that all members of a species evolve together such that each individual organism will produce offspring that are better adapted than the parent was at birth (Shtulman, 2006; Shtulman & Schulz, 2008). While these initial theories continue to play a role throughout the lifespan (Shtulman & Valcarcel, 2012), novel theories are also acquired through everyday experience (e.g., Kempton, 1986; Vosniadou & Brewer, 1994) and through formal education (e.g., Shtulman & Calabi, 2013). We'll refer to such theories as "folk theories," both to differentiate them from full-fledged scientific theories and to avoid the implication that such theories are necessarily "intuitive" in the dual-systems sense.

Folk theories are characterized along three dimensions: structural, functional, and dynamic (Gopnik & Wellman, 2012; Gopnik & Meltzoff, 1997). At a structural level, folk theories specify law-like regularities and involve coherent, abstract, and typically causal representations of the world. At a functional level, theories support important judgments and behaviors, including predictions, explanations, counterfactuals, and interventions. And at a dynamic level, folk theories are revised in light of new evidence. These features differentiate folk theories from other kinds of mental representations, such as heuristics, networks of semantic association, or simple schemas.

The characteristic structural, functional, and dynamic properties of folk theories potentially muddy the distinction between intuitive and deliberative. Some characteristics of folk

theories put them on the "intuitive" end, and it's not a coincidence that they're sometimes called *intuitive* theories: they often generate judgments quickly and are cognitively opaque in the sense that they operate over representations and processes that aren't necessarily explicitly available. They're also often invoked to explain errors. On the other hand, they have some characteristics that align them with deliberation. At a structural level, they tend to involve fairly abstract representations. At a functional level, they support explanations, which involve explicit appeal to theoretical content. And at a functional level, they're responsive to evidence and argumentation – learning processes more naturally associated with deliberation. So how do folk theories fit into a dual-systems approach to the moral domain?

Most approaches to moral psychology recognize an important role for folk theories in analyzing or structuring the *input* to moral judgment. For example, moral judgments can depend critically on causal analyses predicated on folk physical assumptions, and on analyses of an agent's intentions that depend on folk psychological mechanisms (e.g., Cushman, 2008). (Exceptions to this generalization include approaches that deny the existence of folk theories altogether, or that reject the premise that folk scientific analysis "precedes" moral analysis, e.g., Knobe, 2010.) But we wish to suggest something stronger: that folk theories not only structure the input to moral judgment, but can also embody theoretical commitments that play a role in explicit moral deliberation and in "educating" moral intuitions. This can occur in two ways: if the theories themselves contain moral content, or if the theories involve general commitments – such as dualism – that inform and constrain moral judgments. In the next sections, we provide evidence for both of these possibilities.

Evidence for folk moral theories: The building blocks

What might intuitive *moral* theories look like? Such theories can be understood as a special type of folk theory specifically within the moral domain. At a structural level, folk moral theories should be abstract and rule-based in nature. At a functional level, folk moral theories should support moral judgments and justification. And at a dynamic level, moral theories should be responsive to new evidence, be it through direct instruction or more implicit learning mechanisms. We'll review evidence for each of these in turn.

Structurally, there's good evidence that at least some moral "rules" are represented in terms of fairly abstract causal structure (Mikhail, 2011; Waldmann & Dieterich, 2007) and formulated over fairly abstract concepts, even in early childhood (Hamlin, 2013). For example, Waldmann and Dieterich (2007) found that individuals are more willing to accept a utilitarian trade-off that involves harming a few individuals to save a greater number of people if the intervention is targeted at the agent and not the patient. These findings suggest not only that moral "rules" are formulated over abstract causal structure, but that the causal analysis involves morally-relevant distinctions, such as that between agent and patient.

Functionally, folk moral theories should at least partially govern more implicit moral judgments, such as judgments concerning which actions (interventions) are morally permissible. This is what Lombrozo (2009) found in a study investigating whether individuals' explicit utilitarian and deontological moral commitments predict "intuitive" moral responses to trolley car problems. Those participants with explicit utilitarian moral preferences were not only more likely to judge action in trolley car scenarios (all of which involved sacrificing one life for five) more permissible, but also to offer more consistent judgments when two scenarios were presented side-by-side – a manipulation that's been shown to facilitate the extraction and application of rules (Gentner & Medina, 1998).

Finally, moral theories should have dynamic properties – they should change in response to evidence through a process of theory revision. Haidt (2001) influentially suggested that moral intuitions drive moral reasoning "just as surely as a dog wags its tail" (p. 830). But others have suggested that moral intuitions *are* dynamic and open to revision, in perhaps subtle ways. Pizarro and Bloom (2003) proposed that individuals "educate" moral intuitions either through the mere act of thinking, or through selectively exposing themselves to certain experiences in the world. The former mechanism for theory change falls squarely on the side of deliberation: humans can engage in complex courses of private reflection, activating new and sometimes contradictory intuitions. Over time, deliberation of this kind could "tune" intuitions to conform to the outputs of more deliberative reasoning (see also Railton, 2014).

Individuals can also dynamically alter intuitions in more indirect ways – for instance, by controlling their experiences, thus exerting distal control to shift intuitions. Evidence for this comes from work on implicit racial attitudes and the ease with which automatic judgments can be manipulated by a variety of explicit techniques. For instance, participants exposed to positive African American exemplars, both in the lab and in a formal course on racism taught by an African American professor, exhibit reduced implicit biases (Dasgupta & Greenwald, 2001; Rudman, Ashmore, & Gary, 2001), and in light of this, an individual could set out to systematically alter her environment. Studies of this sort illustrate the porous boundaries between intuitive and deliberative processes; intuitions can be tuned up, both in the presence and absence of new experiential data from the world, and although this "learning" process may be initiated by deliberative choice in a distal sense, the learned intuitions could subsequently respond in relatively fast and automatic ways.

Folk moral theories as theories

So far we've provided isolated examples of theory-like structural, functional, and dynamic characteristics within the moral domain. However, such isolated examples are insufficient to support the stronger claim that people possess moral *theories* as such. For this stronger claim, we would want additional evidence that these isolated theory-like elements are integrated into a somewhat coherent whole. For instance, we'd want evidence that the consequentialist commitments that predict trolley car judgments in Lombrozo (2009) are relatively abstract (a structural property) and responsive to evidence (a dynamic property), and that they engage with other relevant moral beliefs.

Systematic coherence is often treated as one of the most compelling sources of evidence for folk theories as distinct from other mental representations. For instance, in the domain of biology, Slaughter and Lyons (2003) taught preschool-aged children about the functional roles of different organs in the body, and found that this impacted their conceptions of death, suggesting a coherent and inter-connected set of biological beliefs related to bodily function. With adults, Shtulman (2006) found that students tended to hold relatively coherent "transformational" or "variational" views of natural selection, rather than clusters of unrelated beliefs. Do we have evidence for such coherence in the moral domain?

Little work has focused on questions of coherence directly, but the study of moral vegetarianism provides an instructive example. At a structural level, we know that moral vegetarianism is supported by relatively abstract beliefs that can be explicitly articulated and applied. Beardworth and Keil (1992), for example, found that moral vegetarians can explicitly identify motivations for their view, citing concerns for animal welfare or a utilitarian concern for environmental sustainability, both of which reflect broad-ranging commitments. At a functional

level, we know that these beliefs guide behavior (i.e., food choices), but also explanations, predictions, and other judgments. Vegetarians, for instance, are more likely to conceive of animals as possessing a wider range of mental states – including the ability to experience pain and suffering (Bastian, Loughnan, Haslam, & Radke, 2012).

And finally, at a dynamic level, there's good reason to believe that beliefs about vegetarianism are susceptible to deliberation and argumentation (e.g., cases of children who become vegetarian independently of their parents, often through discussions with other kids; Hussar & Harris, 2009), but also that explicit commitments to vegetarianism can have long-term effects on an automatic affective response: disgust. Many vegetarians report feeling disgusted at the mere thought of eating meat (Rozin, Markwith, & Stoess, 1997), which would, from an emotivist account, suggest that vegetarians have higher levels of dispositional sensitivity to disgust than non-vegetarians. However, research suggests that disgust does not, at least initially, play a causal role in the decision to become vegetarian: although feelings of disgust toward meat-eating increase over the course of being vegetarian, those who report being motivated by moral concerns (as opposed to health concerns) do not report high dispositional levels of disgust sensitivity (Fessler, Arguello, Mekdara, & Macias, 2003), challenging the idea that affective intuitions are the primary *drivers* of moral judgment. This line of research thus suggests that moral vegetarians' disgust reactions to meat-eating are a byproduct of, as opposed to a cause of, their moral theories. The moral theory has arguably "educated" or "tuned" the individual's affective responses.

The case or moral vegetarianism provides a nice example of how a folk moral theory (in this case about animal rights or welfare) can blur the boundary between intuitive and deliberative processing, with some affective and automatic components, and others that are clearly abstract

and explicit. It's important to note, however, that not all representations with moral content will necessarily conform to the structure of a folk theory. In fact, some findings argue against theory-like representations for some moral content. For example, Goodwin and Darley (2008) found that individuals vary in their meta-ethical commitments to moral objectivism, but that judgments also depend strongly on concrete features of specific moral judgments, such as their content and valence (Goodwin & Darley, 2012). In other words, it could be misleading to classify some people as "moral objectivists" and some as "moral relativists," where the label is taken to reflect an abstractly-represented commitment with broad and systematic scope. Instead, a given individual will appear to be objectivist in some contexts and relativist in others (see also Sarkissian, Park, Tien, Wright, & Knobe, 2011; Uttich, Tsai, & Lombrozo, 2014), suggesting that judgments result from a more contextualized process or representation. More research is needed to truly test the "theory-like" credentials of meta-ethical commitments concerning objectivism, but the example raises an important point: even if some moral judgments result from mental representations that we can properly call theories, it doesn't follow that all do. In fact, it's quite likely that moral judgments are supported by a host of representational formats.

Having considered the case of vegetarianism – which involves moral commitments affecting moral judgment – we now move into evidence from our own work that illustrates a role for high-level *philosophical* commitments in moral judgment.

The case of intuitive dualism

Debates over sanctity of life bioethical issues (e.g., abortion, physician-assisted suicide) often hang critically on the question of when a mere bundle of cells comes to have (or lose) a mind or soul. Bloom (2004; 2006) argues that we're intuitive dualists that separate the physical

body from the non-physical mind; Bloom (2004) argues that "we do not feel as if we are bodies, we feel as if we occupy them" (p. 191), and that this dualist tendency has implications for moral judgment. Greene (2006) similarly explains that "the debate over abortion is ultimately a metaphysical one. The question is not whether a fertilized egg is alive, but whether it is host to a 'human life,' i.e., a human soul. Without a soul in balance, there is no abortion debate. Likewise for the debates over human stem cell research and euthanasia" (p.21).

Dualism has historically been associated with metaphysical commitments about the relationship between the mind and the material. For instance, the Stanford Encyclopedia of Philosophy (2011) defines dualism as "the theory that the mental and the physical – or mind and body or mind and brain – are, in some sense, radically different kinds of thing." Within psychology, however, the term "intuitive dualism" has been used to cast a much wider net. Scales that have been designed to measure intuitive dualist tendencies (Stanovich, 1989) aren't restricted to items that involve the relationship between the mind and the body, but also include concepts related to religious beliefs in a soul or afterlife (e.g., "My consciousness will survive the disintegration of my physical body") and more general views about determinism and reduction in science (e.g., "Knowledge of the mind will forever be beyond the understanding of sciences like physics, neurophysiology, and psychology"). Similarly, experimental manipulations of dualist beliefs (Preston, Ritter, & Hepler, 2013) involve vignettes that also vary in determinism, free will, and reductionist explanations for the human mind.

In recent work (Gottlieb & Lombrozo, in prep), we've created the Dualism+ Scale, designed to measure both narrow metaphysical beliefs related to dualism as well as related but conceptually distinct beliefs, such as those concerning a soul, determinism, scientific reductionism, and epistemological beliefs about the scope of science in explaining mental life

(see Table 2). We find that Dualism+ scores are predictive of five highly controversial bioethical issues that, as Greene (2006) suggested, hinge critically on philosophical commitments: abortion, physician-assisted suicide, cloning humans, cloning animals, and research using embryonic stem cells. However, the component of the Dualism+ scale that drives this relationship is not about the metaphysics of the mind/brain relationship itself ("mind-brain identity"), but about the scope of science and the affective consequences of providing scientific explanations ("scope of science in explaining the mind"): we found that participants' scope of science subscore significantly predicted bioethical judgments, even when controlling for individual differences in political orientation, religiosity, disgust sensitivity, and cognitive style (Frederick, 2005).

Religious commitments to a soul; afterlife beliefs	"Every person has a soul"
Scope of science in explaining the mind	"Explaining everything that makes us human in strictly scientific terms in some way decreases the value of life"
Free will	"People always have the ability to do otherwise"
Determinism	"People's choices and actions must happen precisely the way they do because of the laws of nature and the way things were in the distant past"
Mind-brain identity	"Minds are not the same as brains"

Table 2. The five components of the Dualism+ scale, as supported by a factor analysis, along with a representative item from each.

This finding suggests a causal relationship between commitments concerning the scope of science and bioethical judgments, but it could be that – as with vegetarianism – the

relationship is mediated by affective processes. In fact, bioethical attitudes fall into a class of purity-based sociopolitical issues that are affected by individual differences in disgust sensitivity (Inbar et al., 2009). They are thus canonical examples of moral judgments that are more affective, automatic, and "intuitive" in nature.

Our study also included a measure of disgust sensitivity (Haidt, McCauley, & Rozin, 1994, modified by Olatunji et al., 2007), which revealed that those individuals who are opposed to describing the mind in scientific terms display high levels of disgust sensitivity, even while statistically controlling for political conservatism and religiosity – two factors that have been strongly linked to disgust sensitivity (Inbar et al., 2009; Terrizzi, Shook, & Ventis, 2012). Just as moral vegetarianism can recruit disgust at the thought of eating meat, it could be that beliefs about the scope of science can result in a disgust response to stimuli that implicitly or explicitly violate those commitments, such as physician-assisted suicide. Alternatively, it could be that opposition to scientific descriptions of the mind is itself caused by disgust, which can be elicited by a reminder that the human mind is nothing beyond its physical components (see Rozin, Haidt, & McCauley, 1999 for relevant discussion on animal reminder disgust). Future experimental work – which isolates the direct effect of reductionist descriptions of the mind on state levels of disgust (and visa versa) – will be useful in teasing apart these two pathways.

In sum, our initial findings on the relationship between Dualism+ and views on sanctity of life bioethical issues, such as physician-assisted suicide, suggest a moderate relationship between philosophical commitments and bioethical views. However, the philosophical commitments related to bioethical views concern the scope of science, not the mind-body relationship (narrowly construed). While further research is certainly required, our findings are consistent with the basic proposal that theory-like commitments – in this case relatively broad

and abstract epistemic commitments about science – affect moral judgments, likely in automatic and (in this case) affectively-laden ways. Like the example of vegetarianism – and the examples from the preceding section – this points to the possibility of mental representations that take the form of folk theories that interact with moral judgment, and that crosscut the traditional distinction between intuition and deliberation.

Future Directions

We have argued for the reality of folk moral theories – a form of mental representation distinct from those typically acknowledged by dual-systems approaches. Drawing upon work from cognitive development and cognitive psychology, we suggest that folk theories can involve abstract commitments with structural, functional, and dynamic elements that blur the distinction between intuition and deliberation. In particular, folk theories can involve abstract representations that can be explicitly and deliberatively engaged, but that can also be applied in relatively automatic and implicit ways.

Many elements of our proposal are not in conflict with traditional dual-process approaches. A dual-process theorist can readily accommodate *both* intuitive and deliberative elements to moral judgment, with corresponding mental representations and processes for each. Moreover, such a perspective can accommodate interaction between systems and change over time. In its weakest form, our evidence is merely a warning that the intuition-deliberation distinction can potentially obscure the nature of moral judgment by discounting the contributions of theory-like elements, some of which are intuitive and some of which are deliberative. But in its stronger form, our position argues for the existence of a complex form of representation – a folk theory – that is not merely a collection of elements from an intuitive system and elements

from a deliberative system, but instead a *coherent* representation that does not find a natural home in either system.

The evidence we've marshaled is suggestive, but arguably falls short of establishing this stronger position. In part this is because research has not approached moral judgment with the aim of testing the presence and boundaries of folk theories. Thus many questions remain open, and we see the value of our proposal in part as a spur to further research. We'll conclude by highlighting two directions for such research that we see as especially valuable.

First, one of the most compelling forms of evidence for folk theories – as distinct from other forms of mental representation – comes in the coherence of the mental representation. It's also coherence between intuitive and deliberative elements that arguably poses the greatest challenge to dual-process approaches. But to what extent are folk moral theories coherent? Or to complicate matters further, in what form are they coherent, and is this form of coherence a true challenge to dual-process approaches?

Second, how do education and affect interact in the dynamic tuning of moral judgment? Our evidence for the role of epistemological commitments in bioethical judgment complements Pizarro and Bloom's (2003) proposal that various forms of distal control can revise and reshape the nature of moral intuitions. This view suggests that individuals who oppose scientifically reductionist descriptions of the mind may actively avoid certain forms of education (such as neuroscientific education), or even purposefully engage in religious dialog that argues *against* a reductionist picture of human nature. On the other hand, if these same individuals actively engage with neuroscience, they could experience a dampened emotional response due to a shift in underlying moral intuitions that accrues over time. We could test this hypothesis by gathering longitudinal data on how science education influences metaphysical and epistemological

commitments, both at intuitive and more explicit levels, and therefore impacts bioethical judgments. This empirical question is especially relevant in a culture that is becoming increasingly "scientific" and reductionist in nature (Hook & Farah, 2013). Although Bloom (2004) is skeptical of the extent to which neuroscientific explanations can revise dualist intuitions, conceiving of these commitments as a more general form of folk theory suggests that they may be revised in light of new experiential data.

In sum, we have argued in favor of a unique role for theory-like representations in moral judgment that crosscut the intuitive-deliberative distinction. And although this view is relatively new to the moral psychology literature, it draws heavily upon the literature on folk theories of the natural world, which can be explicit and law-like in principle but engaged in implicit ways. We suggest that this approach is useful in making sense of intuitions regarding fantastical variants on trolley-car dilemmas and other high-conflict scenarios invoked for moral psychology research. But more importantly, and certainly more timely, this approach provides insight into the real-world judgments that divide individuals when it comes to matters of life and death, such as in cases of physician-assisted suicide.

References

Bastian, B., Loughnan, S., Haslam, N., & Radke, H. R. (2012). Don't mind meat? The denial of mind to animals used for human consumption. *Personality and Social Psychology Bulletin, 38*(2), 247-256.

Beardsworth, A., & Keil, T. (1992) The vegetarian option: varieties, conversions, motives and careers. *The Sociological Review 40*(2), 253-293.

Bloom, P. (2004). *Descartes' baby: How the science of child development explains what makes us human*. New York: Basic Books.

Bloom, P. (2004). Natural-Born Dualists. *Edge: The Third Culture*.

Bloom, P. (2006). Seduced by the flickering lights of the brain. *Seed Magazine*, 27.

Carey, S. (2000). Science education as conceptual change. *Journal of Applied Developmental Psychology, 21*(1), 13-19.

Clement, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics, 50*(1), 66-71.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition, 108*(2), 353-380.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment testing three principles of harm. *Psychological Science, 17*(12), 1082-1089.

Dasgupta, N., & Greenwald, A. G. (2001). On the malleability of automatic attitudes: combating automatic prejudice with images of admired and disliked individuals. *Journal of Personality and Social Psychology, 81*(5), 800.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science, 8*(3), 223-241.

Farah, M. J., & Hook, C. J. (2013). The seductive allure of "seductive allure." *Perspectives on Psychological Science, 8*(1), 88-90.

Fessler, D. M., Arguello, A. P., Mekdara, J. M., & Macias, R. (2003). Disgust sensitivity and meat consumption: A test of an emotivist account of moral vegetarianism. *Appetite, 41*(1), 31-41.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, 25-42.

Gentner, D., & Medina, J. (1998). Similarity and the development of rules. *Cognition*, 65, 263–297.

Goodwin, G. P., & Darley, J. M. (2008). The psychology of meta-ethics: Exploring objectivism. *Cognition, 106*(3), 1339-1366.

Goodwin, G. P., & Darley, J. M. (2012). Why are some moral beliefs perceived to be more objective than others? *Journal of Experimental Social Psychology, 48*(1), 250-256.

Gopnik, A., & Wellman, H. M. (2012). Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin, 138*(6), 1085.

Gopnik, A., Meltzoff, A. N., & Bryant, P. (1997). *Words, thoughts, and theories*. Cambridge, MA: Mit Press.

Gottlieb, S. & Lombrozo, T. (in prep). Dissociable components of mind-body dualist attitudes predict bioethical judgments.

Greene, J. D. (2007). The Secret Joke of Kant's Soul. In W. Sinnott-Armstrong (Ed.), *Moral Psychology, Volume 3* (pp. 35-80). Cambridge, MA: MIT Press.

Greene, J. D. (2011). Social neuroscience and the soul's last stand. *Social neuroscience: Toward understanding the underpinnings of the social mind*, 263-273.

Greene, J. D., Morelli, S. A., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2008). Cognitive load selectively interferes with utilitarian moral judgment. *Cognition, 107*(3), 1144-1154.

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

Greene, J. D., Sommerville, R. B., Nystrom, L. E., Darley, J. M., & Cohen, J. D. (2001). An fMRI investigation of emotional engagement in moral judgment. *Science, 293*(5537), 2105-2108.

Haidt, J. (2001). The emotional dog and it rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*, 814–834.

Haidt, J., McCauley, C., & Rozin, P. (1994). Individual differences in sensitivity to disgust: A scale sampling seven domains of disgust elicitors. *Personality and Individual Differences, 16*(5), 701-713.

Hamlin, J. K. (2013). Moral judgment and action in preverbal infants and toddlers: Evidence for an innate moral core. *Current Directions in Psychological Science, 22*(3), 186-193.

Hauser, M., Cushman, F., Young, L., Jin, R., & Mikhail, J. (2007). A dissociation between moral judgments and justifications. *Mind & language, 22*(1), 1-21.

Hussar, K. M., & Harris, P. L. (2010). Children who choose not to eat meat: A study of early moral decision-making. *Social Development, 19*(3), 627-641.

Inbar, Y., Pizarro, D. A., & Bloom, P. (2012). Disgusting smells cause decreased liking of gay men. *Emotion, 12*, 23-27.

Inbar, Y., Pizarro, D. A., Knobe, J., & Bloom, P. (2009). Disgust sensitivity predicts intuitive disapproval of gays. *Emotion, 9*(3), 435.

Inbar, Y., Pizarro, D., Iyer, R., & Haidt, J. (2012). Disgust sensitivity, political conservatism, and voting. *Social Psychological and Personality Science, 3*(5), 537-544.

Keil, F. C. (2011). Science starts early. *Science, 331*(6020), 1022-1023.

Kempton, W. (1986). Two theories of home heat control. *Cognitive Science, 10*(1), 75-90.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences, 33*(04), 315-329.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research* (pp. 151–235). New York: Academic Press.

Lombrozo, T. (2009). The role of moral commitments in moral judgment. *Cognitive Science, 33*(2), 273-286.

McCloskey, M. (1983). Naive theories of motion. *Mental models*, 299-324.

Mendez, M. F., Anderson, E., & Shapira, J. S. (2005). An investigation of moral judgement in frontotemporal dementia. *Cognitive and Behavioral Neurology, 18*(4), 193-197.

Mikhail, J. (2011). *Elements of moral cognition: Rawls' linguistic analogy and the cognitive science of moral and legal judgment*. Cambridge University Press.

Olatunji, B. O., Williams, N. L., Tolin, D. F., Abramowitz, J. S., Sawchuk, C. N., Lohr, J. M., & Elwood, L. S. (2007). The Disgust Scale: item analysis, factor structure, and suggestions for refinement. *Psychological Assessment, 19*(3), 281.

Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science, 36*(1), 163-177.

Perkins, A. M., Leonard, A. M., Weaver, K., Dalton, J. A., Mehta, M. A., Kumari, V., Williams, & Ettinger, U. (2013). A dose of ruthlessness: Interpersonal moral judgment is hardened by the anti-anxiety drug lorazepam. *Journal of Experimental Psychology: General, 142*(3), 612.

Pizarro, D. A., & Bloom, P. (2003). The intelligence of the moral intuitions: A comment on Haidt (2001).

Preston, J. L., Ritter, R. S., & Hepler, J. (2013). Neuroscience and the soul: competing explanations for the human experience. *Cognition, 127*(1), 31-37.

Railton, P. (2014). The affective dog and its rational tale: Intuition and attunement. *Ethics, 124*(4), 813-859.

Robinson, H. (2011). Dualism. *Stanford Encyclopedia of Philosophy*.

Rozin, P., Haidt, J., & McCauley, C. R. (1999). Disgust: The body and soul emotion. *Handbook of cognition and emotion*, 429-445.

Rozin, P., Markwith, M., & Stoess, C. (1997). Moralization and becoming a vegetarian: The transformation of preferences into values and the recruitment of disgust. *Psychological Science, 8*(2), 67-73.

Rudman, L. A., Ashmore, R. D., & Gary, M. L. (2001). "Unlearning" automatic biases: The malleability of implicit stereotypes and prejudice. *Journal of Personality and Social Psychology, 81,* 856–868.

Saad, L. (2010). Four moral issues sharply divide Americans. *Gallup Poll News Service*.

Sarkissian, H., Park, J., Tien, D., Wright, J. C., & Knobe, J. (2011). Folk moral relativism. *Mind & Language, 26*(4), 482-505.

Shtulman, A. (2006). Qualitative differences between naive and scientific theories of evolution. *Cognitive Psychology, 52*, 170-194.

Shtulman, A., & Calabi, P. (2013). Tuition vs. intuition: Effects of instruction on naive theories of evolution. *Merrill-Palmer Quarterly, 59*, 141-167.

Shtulman, A., & Schulz, L. (2008). The relation between essentialist beliefs and evolutionary reasoning. *Cognitive Science, 32*, 1049-1062.

Shtulman, A., & Valcarcel, J. (2012). Scientific knowledge suppresses but does not supplant earlier intuitions. *Cognition, 124*, 209-215.

Slaughter, V., & Lyons, M. (2003). Learning about life and death in early childhood. *Cognitive Psychology, 46*(1), 1-30.

Snyder, L., & Sulmasy, D. P. (2001). Physician-assisted suicide. *Annals of Internal Medicine, 135*(3), 209-216.

Stanovich, K. E. (1989). Implicit philosophies of mind: The dualism scale and its relation to religiosity and belief in extrasensory perception. *The Journal of Psychology, 123*(1), 5-23.

Suter, R. S., & Hertwig, R. (2011). Time and moral judgment. *Cognition, 119*(3), 454-458.

Terrizzi Jr, J. A., Shook, N. J., & Ventis, W. L. (2012). Religious conservatism: An evolutionarily evoked disease-avoidance strategy. *Religion, Brain & Behavior, 2*(2), 105-120.

Trémolière, B., De Neys, W., & Bonnefon, J. F. (2012). Mortality salience and morality: Thinking about death makes people less utilitarian. *Cognition, 124*(3), 379-384.

Turiel, E. (1983). *The development of social knowledge: Morality and convention*. Cambridge, UK: Cambridge University Press.

Uttich, K., Tsai, G., & Lombrozo, T. (2014). Exploring metaethical commitments: Moral objectivity and moral progress. *Advances in Experimental Moral Psychology*, 188.

Vosniadou, S., & Brewer, W. F. (1994). Mental models of the day/night cycle. *Cognitive Science, 18*(1), 123-183.

Waldmann, M. R., & Dieterich, J. H. (2007). Throwing a bomb on a person versus throwing a person on a bomb: Intervention myopia in moral intuitions. *Psychological Science, 18(*3), 247-253.