Introduction

Historically, the study of morality—such as the nature of ‘the good’, what counts as good behavior, what constitutes a good life, etc.—has largely been the domain of philosophers and theologians. But recently psychologists and cognitive scientists have entered the fray and taken on the task of understanding the bases of our moral norms and practices.—for instance, how we arrive at our moral judgments, how we develop our moral concepts, how moral considerations impact other cognition, etc. One line of this cognitive scientific research into morality has pursued the question of whether and how our moral psychological processes might be related to other social cognition, such as mind-attribution. For example, Kurt Gray and Daniel Wegner have recently promoted a fascinating thesis regarding moral perception, which they call the “Moral Typecasting” thesis.¹ A central component of this model is that the cognitive processes involved in seeing something as a moral entity (i.e., a moral agent or a moral patient) are closely intertwined with those processes involved in attributing mental states (i.e., intentionality and phenomenal experience).

The idea that morality and mentality are closely linked is not entirely unique. Philosophers have long associated mindedness with moral considerability,² and the psychological association has also recently been defended by philosophers and cognitive scientists like Joshua Knobe and Jesse Prinz (2008) and Philip Robbins and Anthony Jack (2006). The more provocative aspect of Gray and Wegner’s typecasting thesis is its claim that perceiving a thing as a moral agent prevents perceiving that thing as a moral patient, and vice versa. In what follows, I want both to challenge the conception of morality on which the typecasting thesis is fundamentally based and to raise some concerns with the data offered in favor of moral typecasting. I first argue

that the dyadic definition of morality is far too narrow to fully capture either all of morality or all of moral psychology. Further, even setting aside the problems with the dyadic notion of morality, I hope to show that the experimental data Gray and Wegner appeal to fail to demonstrate the sort of mutual exclusivity of and causal interaction between moral perceptions necessary to support the moral typecasting thesis. Rather, I suggest, the perceptual differences that do show up in the cited studies arguably arise not from a psychological tendency towards moral typecasting, but from confounding features of the characters in the stimuli.

§1. Mind Attribution and Moral Perception

Gray and Wegner propose a model of moral cognition according to which perceptions of certain moral properties are specifically intertwined with attributions of phenomenal mental states (qualitative states, like seeing red or feeling pain), while perceptions of other moral features are distinctively linked with attributions of intentional mental states (e.g., believing and desiring). On this account, when we perceive something to be a moral patient, we instinctively think of that thing as something that is especially prone to experiential mental states; likewise, when we identify an entity as a moral agent, we automatically associate that thing with the capacity to have beliefs, desires, plans of action, and such. This picture is, essentially, composed of three claims. First, moral perception is fundamentally associated with mind perception. Second, just as mind perception has two dimensions, moral perception is dyadic, with one agent and one patient. Third, the two dyads of moral perception—moral agent and moral patient—are inversely related.

In their 2007 paper, Heather Gray, Kurt Gray, and Daniel Wegner report that people attribute mental states along two dimensions: agency and experience. The agency category included characteristics such as having self-control, morality, memory, emotion recognition, planning, communication, and thought. The experience category included characteristics such as the ability to feel hunger, fear, pain, pleasure, rage, desire, pride, embarrassment, and joy, as well as having
personality and consciousness. Gray et al. presented subjects with pairs of entities (e.g., a fetus, an infant, a child, an adult, a robot, God, a chimpanzee, a frog, a dog, a dead woman, and a man in a persistent vegetative state) and asked them which of the two was most likely to have the various characteristics. They found that, for many of the entities, judgments on the two dimensions came apart. God, for instance, “was perceived as having much agency but little experience.”

The basis of the typecasting thesis arises from a correlation the researchers found between the two dimensions of mind perception and the properties of moral agency and patiency. According to Gray et al., “Ratings of deserving punishment for wrongdoing… correlated significantly more strongly with agency than experience, whereas desire to avoid harming… correlated more strongly with experience than agency.” Based on these correlations, the authors conclude that perception of moral agency is linked with the agency dimension of mind perception and that perception of moral patiency is linked with the experience dimension of mind perception. This conclusion is (part of) the basis for the moral dyad thesis (discussed below). Gray and Wegner further reason that, although this two-dimensional analysis of moral perception only entails that moral agency and moral patiency are distinct perceptions, our everyday observations actually suggest something stronger. Namely, they suggest that one kind of moral perception precludes the other, that seeing something as a moral patient prevents us from seeing it as a moral agent (and vice versa), which is the moral typecasting thesis.

While the idea that moral perception is associated with mind perception is relatively straight-forward, both the moral dyad thesis and the moral typecasting thesis contain some ambiguity. The next two sections are committed to getting clear on what, exactly, each thesis is claiming.

---

4 Gray and Wegner (2009), 506.
§2. Clarifying the Moral Dyad Thesis

The starting point for moral typecasting is the idea that all moral perception is dyadic in nature. According to Gray and Wegner:

[I]n a fundamental sense, morality describes a social interaction that takes two—a moral agent who does something right or wrong, and a moral patient who is the recipient of the right or wrong action.\(^5\)

Combine the idea that morality fundamentally involves two entities with the idea that moral perception is essentially linked to mind-attribution, and you get the moral dyad thesis: moral judgments are fundamentally based in our tendency to perceive two minds, with one intentional agent and one suffering patient. Admittedly, there is something intuitively compelling about this thesis. However, the dyadic conception of morality is not without its problems.

First of all, the initial dyadic claim is not explicitly about moral perception or moral psychology, but about morality itself. All of morality. Yet one line of ethical study has, since Aristotle, understood and investigated morality, not in terms of agents and patients, but in terms of how the individual ought to live. Virtue ethics takes morality to be a matter of the individual developing within herself certain character traits, such as wisdom and temperance.\(^6\) And while this question will, undoubtedly, often include considerations of how we should treat and interact with others, most virtue ethicists will deny that such considerations capture the entirety of morality. To reduce all of morality to social interactions between two individuals is to ignore a significant portion of the history of ethical studies.

Of course, given that Gray and Wegner are offering a psychological account of moral perception, it may be more fruitful to understand the idea of the moral dyad as a description of

\(^5\) Ibid, 505.
moral cognition. But even limiting the moral dyad claim to describing moral psychology does not save it from the charge of possibly being too narrow. This conception seems to restrict the entire investigation of moral psychology to situations involving exactly two individuals (no more, no less).

But, this focus seems to ignore a great many of our everyday moral behaviors. Most obviously, it fails to acknowledge that a great many moral exchanges involve more than two individuals. Just consider, for instance, the various ways that corporations, associations, departments, governments, and organizations can behave morally towards their employees, members, customers, citizens, and participants. How are we to fit interactions between individuals and such groups into our moral system? Similarly, it fails to acknowledge the multitude of moral judgments people make every day that do not involve anyone but the individual herself. Consider, for instance, the all-too-common disapproval of homosexuality, or moral opposition to logging practices. In such cases, it is unclear as to who or what, if anything, plays the part of moral patient in such judgments.

Elsewhere (Gray et al. 2012a, b), the moral dyad is described as a cognitive template, used to structure moral perception such that moral situations are interpreted in terms of an intentional agent and a suffering patient, even when no agent (or patient) objectively exists. There they also admit that this template, like most concepts, is somewhat fuzzy and vague, and is not intended to capture every single instance of moral perception. However, they continue, “[t]his dyadic template fits the majority of moral situations because mind perception is as flexible as moral judgment itself.”

---

7 Indeed, this is what Gray et al. (2012a,b) seem to explicitly endorse: “…we suggest that moral judgment is rooted in a cognitive template of two perceived minds—a moral dyad of an intentional agent and a suffering patient.” (101)
8 Gray et al. (2012a), 103.
Clearly, then, the original claim about the dyadic nature of morality is best understood, not as a metaethical claim about the metaphysical essence of morality itself, but as a rough generalization about the nature of moral perception. This is a substantive and interesting claim, worthy of greater scrutiny and consideration than there is room for here. For the sake of argument, I will set aside the question of whether the mind generally and automatically dissects morality dyadically. Instead, my focus will remain on the moral typecasting thesis that there is a fundamental, causal, and inverse relation between perceptions of moral agents and patients.

§3. Clarifying the Moral Typecasting Thesis

According to Gray and Wegner, typecasting can be understood as “[the] tendency to perceive the social world in terms of the two mutually exclusive entities of moral agents and moral patients.” That we see someone or something as a moral patient, on this picture, just seems to make it impossible (or at least difficult) to see them as a full-fledged moral agent, and vice-versa. For instance, Gray and Wegner ask us to consider our attitudes towards the Dickens character Tiny Tim, from *A Christmas Carol*: “It is hard to imagine Tiny Tim,” they say, “as morally responsible for much of anything; the crippled child is a quintessential moral patient, not an agent.” Likewise, they suggest, we have a real difficulty seeing Adolph Hitler or Mohandas Gandhi, prototypical moral agents, as being vulnerable or sensitive (traits Gray and Wegner presume are constitutive of the ‘moral patient’ concept). So, while the 2007 study provides evidence of the perceptual categories being separate and distinct, everyday intuitions (about, e.g., Tiny Tim, Hitler, and Gandhi) suggest that the two categories of moral perception stand in mutual opposition. Rather than the two kinds of perception merely being distinct, moral typecasting presents a psychological picture wherein perception of moral agency and patiency are inversely related. But

---

9 See the special issue of *Psychological Inquiry*, 23(2), for an excellent discussion of the moral dyad thesis.
10 Gray and Wegner (2009), 506 emphasis added.
11 Ibid.
this thesis lends itself to different interpretations, and because there is textual support for both interpretations, it is difficult to determine which interpretation the authors intend.\textsuperscript{12}

The strong interpretation of the thesis—let’s call it the \textit{radical interpretation}—is that perceiving any degree of moral agency entails a complete inability to perceive moral patiency, and vice versa. In their own words, “Seeing someone as a moral agent may preclude viewing them as a moral patient, and seeing someone as a moral patient may preclude viewing them as a moral agent.”\textsuperscript{13} And: “In any moral dyad, then, a person can be either an agent or a patient, not both, and this differentiation into agent or patient within a moral dyad is likely to extend to moral person perception in general.”\textsuperscript{14} These statements suggest that moral typecasting requires the two kinds of moral perception to be mutually exclusive, that we cannot see one thing as both a moral agent and a moral patient.

On the other hand, a less radical and more charitable interpretation—let’s call it the \textit{moderate interpretation}—adheres to the spirit of the thesis, if not the letter. On this interpretation, perceiving something as a moral agent \textit{diminishes} seeing it as a moral patient to some extent (and vice versa). Such a reading is also consistent with claims made in other portions of Gray and Wegner’s paper. For instance:

If moral typecasting influences moral perception, manipulations that enhance moral patiency should \textit{diminish} perceptions of a target’s moral agency, and conversely, a manipulation that enhances moral agency should \textit{undermine} perceptions of the target’s moral patiency… [We] predict moral typecasting: the target who is relatively more of a patient

\textsuperscript{12}Sinnot-Armstrong (2012) raises a similar worry about the moral dyad thesis as presented by Gray, Young, and Waytz (2012a). Gray et al. (2012a, 2012b) explain the moral dyad as a “cognitive template”, which compels people “to understand all immoral acts as dyad and…to perceive an intentional agent and a suffering patient even when they may be objectively lacking.” (2012b, 207-8.) Importantly, this description does little to clarify the degree of mutual exclusivity moral typecasting imposes on agent- and patient-perception.

\textsuperscript{13}Gray and Wegner (2009), 506.

\textsuperscript{14}Ibid, 507.
should be seen as relatively less of an agent, and the target who is relatively more of an agent should be seen as relatively less of a patient.\footnote{Ibid, 511, emphasis added.}

Although it is not always obvious which interpretation of moral typecasting the authors intend, only the moderate version is plausible in light of the authors’ experimental data.

First of all, the radical interpretation seems unlikely, given what Gray et al. (2007) say about mind perception. The authors in the original article note that, for some entities (e.g., adult humans), the \textit{agency} and \textit{experience} dimensions are strongly correlated. So, if moral typecasting takes the association between the two kinds of moral perception and the two dimensions of mind perception to be fundamental, then the fact that mental dimensions correlate for certain entities would undermine the thesis that the two kinds of moral perception are entirely mutually exclusive. That is, insofar as mind-attribution judgments about \textit{agency} and \textit{experience} are linked with perception of moral agency and perception of moral patiency, respectively, and if judgments about \textit{agency} and \textit{experience} for adult humans are strongly correlated, then perception of moral agency and perception of moral patiency in human adults must be likewise positively correlated. But, then, there would seem to be a conflict between the data showing that dimensions of mind perception are not always mutually exclusive and the radical interpretation of moral typecasting.

Moreover, as will be seen in later sections, the new data offered in support of typecasting does not suggest anything like mutual exclusivity in the radical sense. Rather, the data are purported to show that those who score highly on one kind of moral perception score less highly on the other. That is, Gray and Wegner interpret (certain of) their data as showing not that an individual perceived to be a moral agent is attributed \textit{no capacity for pain whatsoever}, but that such an individual is attributed \textit{less susceptibility to pain} than an individual perceived to be more of a moral patient.
As such, a charitable analysis of the typecasting thesis should rely on the moderate interpretation, wherein one kind of perception diminishes but does not wholly preclude the other. The rest of this paper will focus primarily on that interpretation. While this version may seem preferable, even reasonable, when compared to the radical interpretation, I do not think it manages to avoid a significant problem: namely, it fails to be supported by the experimental data. That certain entities receive a comparatively higher rating for agency while also receiving a comparatively lower rating for patiency is not itself proof of an inverse, causal relation between the two kinds of moral perception. A thing’s being perceived as highly agentive may well correlate with that thing’s being perceived as less-than-fully patient-like without the one being the cause of the other. In order to support the typecasting thesis, the experiments must demonstrate that such a causal connection in fact exists. I argue, below, that they do not. Indeed, the data lend themselves to alternative explanations that strike me as both more obvious and more intuitively plausible than moral typecasting.

§4. Analyzing the Studies

In nearly all of the vignettes used to support moral typecasting, I think it is possible to identify confounding variables that allow for alternative explanations of the results. For instance, one study found that people tended to rate a normal adult man as more responsible for his action and as having acted more intentionally compared to a five-year-old boy performing the same action. That study also found that participants judged the adult to be less capable of feeling pain than the child. But these comparative differences do not necessarily point to an inverse relation between perceptions of moral agency and moral patiency. The fact that normal adults differ from five-year-olds in terms of perceived agency and patiency just makes sense, not because of moral typecasting, but because of our conception of adults and our conception of children.

16 In this case, knocking over a tray of glasses.
The apparent inversion of comparative differences, I argue, is driven entirely by the fact that normal adults are conceived of as possessing the characteristics central to agency, while children are conceived both as lacking some of those characteristics and as possessing more of the characteristics related to patiency. But this does not show that those perceptions are inversely related; it only shows that different entities possess the relevant faculties in differing quantities or degrees.

Given that the measures of perceived agency took the form of questions about being responsible for one’s actions and about acting intentionally, it’s not surprising that children scored relatively low. They lack the psychological sophistication (foresight, moral knowledge, etc.) that we typically require in order to hold someone fully morally responsible; the sort of sophistication that we typically assume normal adults possess. And given that the measure of perceived patiency was the ability to feel pain, it’s also not surprising that children score relatively high. While children’s pain may not hurt any more than adults’, children haven’t developed the mechanisms for coping with pain that most adults have. But neither of these facts has anything to do with moral typecasting per se. These differential attributions, I suggest, are merely part of our conceptions of children qua children, not the product of seeing them as moral patients.

Arguably, the same sort of argument can be run, with some variation, against all of the moral typecasting studies. And the possibility that differences in judgments could be caused by differences in the targets’ characteristics places a burden of proof on typecasting’s supporters to show that there is a genuine interaction between the two kinds of moral perception. In the remainder of this section, I attempt to show that they fail to meet this burden. I intend to show that the differences in agency and patiency ratings are the consequence, not of moral typecasting, but

---

17 Hence the popular tendency to accuse someone of “acting like a baby” when fussing excessively over a mild injury. We commonly expect adults to ignore a certain amount of pain, to “soldier on” or “man up,” but we do not typically expect the same from (5-year-old!) children.
of features in the vignettes that are orthogonal to perceptions of agency and patiency. In doing so, I want to break Gray and Wegner’s studies into two separate categories. The first contains those studies that relied on targets that clearly differed from each other in ways beyond mere agency and patiency; the second contains those studies in which the targets were roughly physically and psychologically equivalent but behaved differently (to some degree) and in a relevant way.\textsuperscript{18}

§4.1 Fundamentally Different Entities

In the first category, which includes roughly two-thirds of Gray and Wegner’s vignettes (1a, 1b, 2, 3a, 4b, 6, and 7), the roles of ‘agent’ and ‘patient’ were filled by distinct sorts of individuals, i.e., by individuals explicitly described as possessing different properties. Studies 1a and 1b, for instance, compare an average 25 year-old male with a 5 year old and a mentally challenged adult, respectively. The participants in studies 2 and 7 made judgments about the Dalai Lama, Mother Teresa, Ted Bundy, a high school teacher, a radiology technician, an orphan, and a rape victim.\textsuperscript{19} In study 3a, participants were presented with two men, one of whom is genetically more sensitive to pain, and one who is genetically less sensitive to pain. Study 4b compared a psychopath working in a nursing home with a corporate CEO. Finally, study 6 compares an ex-British Army pilot with an 11 year old orphan. Even without describing the various measures, one can descry a pattern in the experimental targets: individuals intended to be perceived as agents differ from patients in many important ways, not just in virtue of being a moral agent.

\textsuperscript{18} A brief caveat: the fact that Gray and Wegner report data from several studies makes the present task somewhat difficult. I will not attempt to discuss all seven experiments in the current paper, but will proceed on the assumption that the experiments discussed are representative samples, in which case the criticisms offered against the specific experiments will generalize and apply with equal force to the rest of the studies.

\textsuperscript{19} Study 2 also included Hitler, Gandhi, Martin Luther King, Jr., Osama bin Laden, Britney Spears, Prince Harry, and a man with Down’s Syndrome.
I have already offered some reasons for why people attribute intentionality, responsibility, and pain-capabilities differentially between normal adults and children, reasons that are not about moral typecasting. This reasoning applies with relatively little alteration to the other type-casting studies, which involve a mentally challenged adult, an orphaned 11-year old, a psychopath, and an individual with a greater genetic sensitivity to pain. In all of these studies, the comparison is between entities that differ in cognitive and physical sophistication. And these differences explain why participants would rate the entities differently in terms of acting intentionally and responsibly and in terms of capacity to experience or endure pain. There simply is no need to posit moral typecasting to explain these data.

Take, for instance, the study in which participants were given vignettes about Michael, “the CEO of a large software company,” and Jeffrey, who “works in a nursing home, but is secretly a psychopath.” Participants were asked which of the two was “more powerful, more determined, more daring, and more dominant,” and which was more capable of evil. These measures were taken to indicate how much of an agent simpliciter and how much of a moral agent each was perceived to be. They found that Michael was considered significantly more powerful, determined, daring, and dominant than was Jeffrey; they also found that Jeffrey was considered more capable of evil than Michael. Thus, Michael was seen as more of an agent than Jeffrey, while Jeffrey was seen as more of a moral agent. Gray and Wegner then told participants that “Michael and Jeffrey are both out walking one night when there is an evil murderer on the loose,” and assessed perceptions of moral patiency by asking which of the two was more likely to be a victim of the murderer. Unsurprisingly, perhaps, participants said that Michael was

\[20\] Importantly, this does not rule out Jeffrey being perceived as an agent himself; it only shows that he is not perceived to be as powerful, determined, daring, and dominant as Michael. Gray and Wegner never justify the reduction of agency to these particular properties, even though it is philosophically robust to claim that they constitute agency simpliciter.
more likely to become a victim, which is taken as evidence of an interaction between perceived moral agency and perceived moral patiency.

However, it seems entirely reasonable that people judged Jeffrey as less likely to be a victim in virtue his being a psychopath, who “enjoys nothing more than hurting other people [and] takes special glee in hurting animals.” There are all sorts of reasons, aside from his moral agency, why Jeffrey would be seen as better able to fend off an attacker: he is obviously well-practiced at inflicting harm and would likely not hesitate to do so should he be attacked; his experience at inflicting harm reasonably makes him more adept at recognizing the dangerous situation; or maybe people just think that psychopaths are better at recognizing other evil people, making Jeffrey more capable of identifying the would-be attacker. The potential explanations are numerous and, importantly, do not rely on a process of moral typecasting so much as the simple recognition of relevant implications of Jeffrey’s psychopathy.

Arguably, then, the majority of the apparent evidence for typecasting can be chalked up to features of the characters in their vignettes—features besides perceived moral agency and patiency—and do not definitively support even a moderate version of the typecasting thesis. The differences between judgments can largely be explained by confounding features of our conceptions of children, celebrities, royalty, and orphans, and of monks and murderers, independently of their perceived moral agency.

---

21 Ibid, 514.
22 The remaining studies in the first category are susceptible to a more specific version of the present challenge. In those studies, participants were presented with particular individuals, such as the Dalai Lama, Mother Teresa, Martin Luther King, Jr., Hitler, Ted Bundy, Osama bin Laden, Britney Spears, Prince Harry, a radiology technician, a high school teacher, an orphan, a man with Down’s Syndrome, and a rape victim. Again, there appeared to be a negative correlation between judgments of moral agency and judgments of moral patiency.

But, again, this should not be surprising given (a) the assessments of perceived agency/patiency and (b) the different features possessed by the various targets. In study 2, for instance, they assessed perceived patiency by asking participants how much pain each individual would feel were they to accidentally cut their foot on a piece of glass, and how easy it would be to take advantage of this person. Given the previous discussion about common conceptions of children, it should not be surprising that they found the orphan to be rated extremely high on experienced pain. Likewise, given our conceptions of celebrities and royalty as living relatively plush and extravagant lifestyles, we ought not be surprised that they might be seen as unaccustomed to (and, thus, experience more acutely) personal injury. Nor should we be surprised to find that the Dalai Lama (a monk whose life’s work is to train the body and mind to overcome pain and suffering) and Ted Bundy (a psychopathic serial killer) were judged to experience relatively little pain. But we need not appeal to a moral typecasting thesis to explain such differences. We can explain such differences by our conceptions of celebrities, royalty, and orphans, and of monks and murderers, independently of their perceived moral agency.
tions of the entities and individuals involved, so the fact that attributions of moral agency are negatively correlated to attributions of moral patiency should not be understood as evidence of a direct, causal interaction between the two. The next section turns to the remaining studies used to argue for moral typecasting.

§4.2 Importantly Different Characteristics

The second category of vignettes involves targets that are presumed to be physically and psychologically similar; in these vignettes, distinguishing agent from patient occurred via the characters displaying different behaviors. In a study that is fairly representative of the category, the vignette describes Michael and Jeffrey, who are both employees (presumably of equal age, ability, status, etc.) at an eco-friendly company that has recently been taken over by a CEO who intends to scrap the eco-friendly policies in order to increase profits. According to the story, Jeffrey takes a stand against the CEO by walking out of the meeting, after which Michael does the same.

Presumably, in virtue of being the first to walk out, Jeffrey is perceived as more of a moral agent than Michael, while Michael—merely following Jeffrey’s lead—is perceived as less of a moral agent. The researchers verified this by asking participants to choose between Jeffrey and Michael as to who deserved more praise for their action. Perhaps unsurprisingly, participants chose Jeffrey. The moral typecasting thesis predicts that Jeffrey should then be seen as less of a moral patient and, thus, would be seen to feel less pain than Michael. To test this prediction, participants were told that Michael and Jeffrey had scalding coffee spilled on them while dining at a restaurant, then asked to choose whether Jeffrey or Michael felt more pain. Participants

---

23 Of course, just because Jeffrey deserves more praise, it does not follow that he is, in fact, perceived to be more morally agentive. Even if we define a moral agent as an entity that can earn praise or blame, it does not follow that (degree or amount of) moral agency is the only factor that determines an entity’s praiseworthiness. It may well be that Jeffrey was deemed more praiseworthy, not for being relatively more morally agentive but for acting first.
tended to respond that Michael felt more pain than Jeffrey, which seems to confirm the typecasting thesis.

However, it is altogether possible that subjects in this study took the behavioral differences between Michael and Jeffrey as evidence of differences in character that are not specifically about moral agency and paity. It may be that the behavioral differences between the targets in these vignettes provided participants insight into the targets’ different psychologies, and that these psychological differences influenced participants’ attributions of intentionality, responsibility, vulnerability, etc. Thus, we might interpret any differences in ratings of perceived patiency as consequences of participants perceiving different, relevant personality traits in one of the target characters, and not as a consequence of perceived agency.

My criticism here parallels the one raised in the previous section: just as Jeffrey’s psychopathy provides a plausible alternative explanation for his being seen as less vulnerable than Michael, the difference in perceived patiency in the present study may plausibly be explained by whatever character trait of Jeffrey’s subjects took to be the reason for his being the first to take a stand against the CEO. For instance, Jeffrey’s walking out may suggest to participants that he is not hampered by the discomfort or anxiety or fear of punishment by his boss, or at least is less hampered than is Michael. Or, perhaps, he is seen as having the capacity to overcome those trepidations. Neither of these counts as being more morally agentive, yet either would be a plausible explanation of why Jeffrey would be viewed as slightly less susceptible to the pain of scalding coffee: he simply is not easily discomforted, compared to Michael, or he just has a greater capacity to overcome such negative experiences.

Given that the difference in perceived patiency is plausibly explained by something other than the difference in perceived agency, the data do not obviously and exclusively support the moral typecasting thesis. This reasoning, I suggest, applies with equal force to the remaining ex-
experiments offered in support of typecasting. As such, unless one already accepts the moral typecasting thesis, there is no independent reason for ruling out the interaction between perceptions of non-moral characteristics and attributions of sensual susceptibility.

§5. General Discussion

Throughout the paper, I have offered reasons for thinking that the moral typecasting studies fail to show what they attempt to show. In so doing, a theme has emerged; namely, the observed differences in dependent variables can plausibly be explained by something other than moral typecasting, so the data do not exclusively support the moral typecasting thesis. The reasoning employed in its favor ignores other potential factors—beyond perceived moral agency and patience—that may lead subjects to attribute certain targets increased intentionality, responsibility, and pain-sensitivity. It is entirely possible, if not altogether more plausible, that such attributions are based on subjects picking up on other, non-moral traits in the vignette’s characters. What is taken as proof of perceived moral patience may just as well be the product of perceived weakness, which needn’t have any psychological connection to moral patience. It is also a mistake, I think, to reason that since moral agents deserve praise/blame, it must follow that attributing to a thing more praise or blame equals ascribing that thing more moral agency (*mutatis mutandis* for moral patience). This ignores all of the other factors that go into ascribing praise and blame, such as one’s beliefs or desires or alternative courses of action.

My arguments against typecasting turn largely on the idea that there are alternative explanations for the data presented in favor of moral typecasting. In general, I think that the measures used in these studies are tracking more than just moral agency and patience, which leaves open the possibility that the observed differences are caused by something other than an inverse relation between the two kinds of moral perception. Specifically, I’ve suggested that the
various differences reported as evidence of typecasting can be explained by various features of the characters in their vignettes, features aside from perceived moral agency or patiency.

However, one may respond to this objection by suggesting that, while the force of each of the studies might be undermined, there nonetheless remains a general trend that lends prima facie plausibility to the moral typecasting thesis. It’s the overall trend, they might say, that evinces the inverse relation between perceptions of moral agency and patiency.

But this response ignores the more general objection that effectively undermines each of the experiments. Since each of the data points contributing to the overall trend is based in fallacious logic and/or includes potential confounding variables, one cannot draw legitimate general conclusions from the larger collection. A general trend only holds if the various data points on which it is based are generally accurate. But since there are potential confounds and/or alternate explanations for the majority of typecasting experiments, any generalization drawn from those individual points will be similarly flawed.

A second response may appeal to a kind of cognitive parsimony: in explaining away participants’ responses in the typecasting studies, I have to appeal to numerous conceptual entities (e.g., adult, child, psychopath, etc.), yet typecasting needs only invoke the two relatively simple categories of moral agent and moral patient. In this way, it may seem that the typecasting thesis offers a simple yet elegant explanation for a variety of moral psychological intuitions. Judgments of responsibility, intentionality, and so on fall out of perceiving an entity either as an agent or as a patient, which is simpler than searching through all of the various potential concepts that might apply to the individual.24

It is not entirely clear, however, that this picture actually depicts greater cognitive parsimony. Indeed, typecasting seems to require an additional cognitive apparatus. Given that we

---

24 Thanks to Adam Shriver for raising this point in his comments at the 2012 meeting of the Southern Society for Philosophy and Psychology.
already possess concepts of (or beliefs about) adults and children and psychopaths and so on, there is no conceptual proliferation in my explanation. I am simply suggesting that we can appeal to our already-existent understanding of the world and its inhabitants—via social schemas, folk theories, or whatever—to assign differing degrees of moral responsibility, pain-susceptibility, etc. It seems an unnecessary additional step to route those judgments through a *moral typecasting* mechanism (or process or whatever), since we are already perceiving such entities as ‘normal adult’ or ‘child’ or ‘psychopath’. Thus the explanation I offer is actually more parsimonious than that being offered by Gray and Wegner.

Perhaps, though, the concern in this objection is not about a conceptual or cognitive-procedural parsimony, but rather suggests a kind of explanatory parsimony: typecasting offers a single explanation for all of the data across Gray and Wegner’s studies. While I must invoke a different explanation for each of the results, typecasting offers a single, unified theory capable of explaining all the data.

This challenge seems to presume that there is, in fact, a single, recurrent phenomenon in need of explanation. If the same process is producing all of the judgments (concerning responsibility, etc.) across all of Gray and Wegner’s experiments, then it would certainly be beneficial and preferable to have a single explanation for that process. But it is only a virtue of explanatory parsimony if the various phenomena are, in fact, all tokens of the same type of event. If I am right, and people are arriving at responsibility and pain-susceptibility judgments via different cognitive pathways, then unifying them all under one and the same process would actually be misleading. In order for the challenge of explanatory parsimony to be successful, it would have to be the case that people are, in fact, typecasting. But my objection offers some reason to doubt that the data sufficiently evince typecasting, and to dismiss this objection on the grounds that typecasting (or some singular phenomenon) is occurring is to beg the very issue in question.
Finally, one might reply to my general objection that the features I point to as confounding variables are actually constitutive of the concepts ‘moral agent’ and ‘moral patient’ at play in the typecasting hypothesis. The (presumably) non-moral characteristics I appeal to for potential alternative explanations of the differences in, e.g., pain attribution are simply part of what it is to be a moral agent/patient. The psychological and physical differences I point to as fundamental differences between entities just are the features that cause us to typecast something as a moral agent or a moral patient. As such, I’m just describing moral typecasting in different terms. But then my objections to typecasting are not really objections at all, as the other features are really just components of moral agency/patiency. Thus, when I point out, for instance, that Jeffrey’s being a psychopath (in study 4b) is the real factor causing subjects to rate him as less vulnerable to assault than Michael, I’m actually exemplifying moral typecasting behavior.

This reply, though, threatens to deflate the moral typecasting thesis to the point of insignificance. If it turns out that any feature that might cause us to see an entity as less vulnerable or less sensitive to pain constitutes moral agency, and if any feature that causes us to see someone as less worthy of praise or blame just is a feature of moral patiency, then moral typecasting is in danger of acquiring tautological status. That is, if we allow that the features I take to be responsible for differences in attributions of pain and praise-/blameworthiness (e.g., toughness vs. weakness, background knowledge vs. ignorance) are all just part of our ‘moral agent’ and ‘moral patient’ concepts, then the moral typecasting thesis is reduced to claiming that seeing a thing as being relatively tough diminishes seeing that thing as being relative weak. If typecasting something as a moral agent just means attributing it a cluster of features and typecasting as a moral patient just means attributing the opposing features, then all that the typecasting thesis amounts to is that we don’t attribute directly-contradictory properties to a particular entity.
But clearly this is not the thesis that Gray and Wegner intend to defend. Rather, they take themselves to be presenting a novel and provocative picture of moral cognition. As such, I think the above response is not available to them, since composing the concepts of ‘moral agent’ and ‘moral patient’ out of any and all of the features pointed to in my original objections would reduce the moral typecasting thesis to a rather mundane and uninformative theory.

**Conclusion**

The study of moral cognition is important to advancing our understanding of humanity, as so much of our lives are concerned with reasoning about what we (and others) should and should not do. Such knowledge can also shape our philosophical understanding of good and bad moral judgments (since *ought* implies *can*, philosophical theories of how we ought to form moral judgments are constrained by what our cognitive system is capable of doing). Thus, it is of some importance that our empirical investigations stand up to rigorous evaluation. Sadly, the data offered in favor of moral typecasting does not withstand such an evaluation. As provocative and entertaining as the thesis is, the reasoning behind it and the evidence for it are imperfect. In the end, more empirical work needs to be carried out to demonstrate that moral perception, in fact, involves typecasting.
References


nology and the Cognitive Sciences, 7(1), 67-83.

