PHILOSOPHICAL PERSPECTIVES

Philosophical Perspectives, 27, Philosophy of Language, 2013

GENERICS, GENERALISM, AND REFLECTIVE EQUILIBRIUM: IMPLICATIONS FOR MORAL THEORIZING FROM THE STUDY OF LANGUAGE*

Adam Lerner
Princeton University

Sarah-Jane Leslie
Princeton University[1]

Ethicists often find themselves facing what we call "familiar epistemic dilemmas." They want to endorse the moral judgment that P when they reflect on a particular case, but they also want to endorse a general principle that implies that P is false. In conflicts of this sort, they must either reject their judgment about the case in favor of the judgment implied by the principle, or else revise their principle so that it conforms with their judgment about the case. *Generalists* in moral epistemology believe that when these conflicts arise, the best default strategy is to favor our general principles over our judgments of particular cases, whereas *particularists* privilege judgments about particular cases. In this paper, we offer a novel argument against generalism.[2]

Recent work in cognitive science and psycholinguistics suggests that people often endorse false universal generalizations (e.g., "All ducks lay eggs") when there are nearby generic generalizations that are true (e.g., "Ducks lay eggs"). In this paper, we argue that this "generic overgeneralization effect" extends to the moral domain, and that this psychological fact undercuts prominent philosophical justifications for generalism. In brief, the generic overgeneralization effect provides a debunking explanation of people's intuitive attraction to

*We are indebted to Sara Aronowitz, Daniel Bernston, Matthias Brinkmann, Mara Bollard, Paul Boswell, Daniel Drucker, Christopher Freiman, Robbie Hirsch, Benedikt Kahmen, William MacAskill, Nils-Hennes Stear, Eric Hubble, Zoë Johnson-King, Mark Johnston, Hrishikesh Joshi, Sydney Keough, Barry McGuire, Chip Sebens, Russ Shafer-Landau, Phillip Stratton-Lake, Nina Windgätter, Daniel Wodak, Jack Woods, Richard Yetter-Chappell, Robin Zheng, and audiences at Oxford, the Central European University, and the University of Michigan for helpful feedback and discussion.

general moral principles by defeating any non-inferential justification generalists take themselves to have for endorsing such general moral principles.

In section 1, we review various justifications for endorsing generalism in moral epistemology. In section 1, we distinguish between types of moral principles that generate familiar dilemmas (e.g., universal generalizations) and types that don't (e.g., generic generalizations). In section 2, we review evidence in favor of the generic overgeneralization effect and argue for a particular account of how people process and evaluate universal generalizations. In section 3, we argue for the existence of a moral generic overgeneralization effect, according to which people endorse universal moral generalizations that are false because they mistake them for generic moral generalizations that are true. We go on to argue in section 4 that the moral generic overgeneralization effect undermines some of the most prominent justifications for generalism in moral epistemology. In section 5, we consider objections to the arguments from the previous section. We conclude in section 6 by sketching the conditions under which we should continue to rely on general moral principles when faced with familiar dilemmas.

Section 1: Why Generalism?

Philosophers often find that some of the moral principles they find most plausible conflict with some of their most strongly held judgments about particular cases. When this happens, philosophers sometimes revise their principle in order to accommodate their judgment about the particular case. Call this "taking the particularist route." Other times, philosophers revise their judgment about the particular case in order to bring it into accord with their principle.[3] Call this "taking the generalist route."

Philosophers offer different reasons for taking the generalist route. One prominent strategy consists in casting special doubt on our judgments about particular cases. For instance, in response to the particularist demand to build moral theories up from our judgments about particular cases, Peter Singer asks, "Why should we not rather make the opposite assumption, that all particular moral judgments we intuitively make are likely to derive from discarded religious systems, from warped views of sex and bodily functions, or from customs necessary for the survival of the group in social and economic circumstances that now lie in the distant past? In which case, it would be best to forget all about our particular moral judgments, and start again from as near as we can get to self-evident moral axioms." (1974, p. 516)

A less radical strategy consists not in finding some special reason to discount our judgments about particular cases, but in finding some special reason to privilege our intuitions about general principles. In other words, the goal of this strategy is to find some positive epistemic status that our beliefs about moral principles enjoy but that our beliefs about particular cases lack. As Singer suggests, one way to do this is to claim that the general moral principles we find plausible are self-evident. Although the details get spelled out differently on different theories, modern defenders of self-evidence theory all agree that propositions are self-evident when they can be justifiably believed solely in virtue of being adequately understood. For instance, Robert Audi tells us that "an adequate understanding of a self-evident proposition can enable one to see its truth by virtue of apprehending conceptual relations and without relying on premises." (2009, p. 188) Russ Shafer-Landau claims that "an adequate understanding of the content of self-evident beliefs can be enough to constitute justification, whether or not we know or suspect that we possess such justification." (2003, p. 250)

Although it is open to self-evidence theorists to hold that propositions at any level of generality may be self-evident, self-evidence theorists need not believe that all moral judgments are self-evident, and many reject the idea that moral judgments about particular cases can be self-evident. In this way, self-evidence theorists can dismiss particular moral judgments that are not self-evident when they conflict with general moral principles that are self-evident. As Shafer-Landau suggests, we may justifiably come to believe that "holding a self-evident belief is sufficient to license one's rejection of all competing beliefs."[4] (2003, p. 250)

A related but distinct way to privilege judgments about moral principles over particular cases is by way of *seeming state theory*.[5] The most popular version of seeming state theory comes from Michael Huemer. According to Huemer, the mere fact that a proposition seems true can be enough to make us justified in believing it: "If it seems to S that *p*, then, in the absence of defeaters, S thereby has at least some degree of justification for believing that *p*."[6] (Huemer, 2007, p. 30) When it *seems* to us that a general moral principle is true, this may provide us with an additional reason to believe it over and above any other reasons we may have to believe it, and this may leave us more justified in retaining it than in adopting an incompatible judgment about the particular case (e.g., if our judgment about the particular case is not accompanied by an analogous seeming state).

The final (and perhaps most common) reason to privilege our general moral beliefs over apparent counterexamples is that we are more *confident* that they are true than we are that our particular judgments are true, or that we are more *attached* to them, or that we find them more *entrenched* in our belief system. This way of thinking of what justifies an epistemic move emerges out of the literature on the method of reflective equilibrium. Describing the method as applied to our practices of inductive inference, Nelson Goodman writes, "A rule is amended it if it yields an inference *we are unwilling* to accept; an inference is rejected if it violates a rule *we are unwilling* to amend." (1954, p. 64, emphasis added) Rawls tells us that a moral principle is reasonable only "when a subclass of considered judgments, rather than the principle, *is felt to be mistaken* when the principle fails to explicate it." (1951, p. 188, emphasis added) Considering a similar conflict between general background theories and ("level I") judgments about lower-level principles and particular cases, Norman Daniels argues that we should typically revise a general theory only if "we can give [it] up *more easily* than we can accept the new level I judgment." (1979, pp. 266–267, emphasis added) On this way of thinking, the justification for taking the generalist route in response to familiar dilemmas derives from the (supposed) fact that we typically are more attached to, or more confident of, our general beliefs.

(It is important to note that there is a similar seeming, but in fact quite distinct, justification to which a generalist might appeal. That is, privileging a general principle over a class of particular judgments may be justified when and because those principles better unify and explain the entire set of particular judgments than any competing principle. For this justification to be plausible, generalists would have to provide some reason to think that a principle that is incompatible with some of our particular judgments could be better than a related, *more restricted* or *exception-tolerating* principle that is amended to accommodate our intuitions about these particular cases. We submit that, in the absence of some independent reason to doubt our judgments about these particular cases, we should always favor the principle that can accommodate more of our judgments. This, however, is consistent with epistemological particularism. For this reason, we set aside this inferential justification in the ensuing discussion, since it cannot in itself provide reason to hold on to universal principles in the face of apparent counterexamples.)

While this brief overview of possible justifications for generalism is far from exhaustive, it features the most prominent strategies on offer. Furthermore, with the exception of Singer's debunking strategy, all of the views mentioned here share one critical feature in common: They hold that our beliefs about general moral principles can be *non-inferentially* justified. Whether we are non-inferentially justified because the principles are self-evident, or because they seem to be true, or merely because we find ourselves especially confident in them, these views all entail that our beliefs about general moral principles can enjoy some positive epistemic status that is independent of their relationship to our judgments about particular cases or any of our other beliefs. This shared feature of the views renders them susceptible to the sorts of objections we level against them in section 4.[7]

Section 2: Which Kinds of Moral Principles Give Rise to Familiar Epistemic Dilemmas?

Moral principles come in a variety of flavors, varying on a number of dimensions. One dimension on which moral principles vary is their *scope*, or *the domain over which they quantify*. The kinds of moral principles that generate familiar dilemmas are those that express *universal* generalizations. Universal generalizations are statements of the form, "All acts of lying are wrong," "Lying is always wrong," and "If an act is an act of lying, then it is wrong." Because universal generalizations ascribe a particular property to every member of a kind or category, they are true when every member of that kind has that property and false when there is at least one member of that kind that lacks that property. The universal generalization, "Lying is always wrong," is commonly regarded as false, since there seem to be some possible acts of lying that are not wrong. For instance, if lying to the would-be murderer about the whereabouts of his would-be victim would not be wrong, then "Lying is always wrong" is false.

Some moral principles, however, express generalizations that can remain true even if there are some members of the kind that lack the property in question. These sorts of principles subdivide into two types. Principles of the first type are themselves universal generalizations, but ones that explicitly specify a set of conditions under which the generalization over the kind fails to hold. For instance, in light of the observation above, one might endorse the moral principle, "Lying is always wrong, except when doing so would save an innocent person's life." Because these principles explicitly enumerate the conditions under which the generalization fails to hold, it is always possible to determine whether such a principle conflicts with a judgment about a particular case. It is for this reason that these restricted universals are capable of giving rise to familiar epistemic dilemmas.

Alternatively, however, sometimes moral principles do not explicitly specify the conditions under which the generalization fails to hold. Instead, these principles merely gesture at the conditions under which we should expect the generalizations not to hold. Principles of this sort come in a variety of guises: "*All else being equal*, lying is wrong," "*In privileged conditions*, lying is wrong,"[8] "*Normally,* lying is wrong," "*Usually*, lying is wrong," "*As a rule*, lying is wrong," and even just "Lying is wrong." Each of these principles can remain true in the face of exceptions. We might believe that there was nothing wrong with Germans lying to Nazis about the fact that they were hiding Jews in their houses, but these principles could remain true all the same.

Although these kinds of principles all tolerate exceptions, the reasons for their tolerance vary. Some limit the domain of the generalizations to "privileged" or "normal" conditions, remaining true so long as those members of the kind that lack the ascribed property are in unprivileged or abnormal conditions. Others express *statistical* generalizations whose truth requires only that some percentage of members of the kind have the ascribed property. Others express *generic* generalizations, whose ability to remain true in the face of exceptions resists explanation in terms of statistical facts or domain restriction. For instance, generic generalizations like "Cardinals are bright red," and "Snakes are venomous" can remain true even if less than half of all cardinals (i.e., the adult males) are bright red and less than half of all snakes (and only 15% of snake species) are venomous.[9] Likewise, it is in no sense normal for a cardinal to be male or a snake to be venomous, and yet "Cardinals are bright red" and "Snakes are venomous" are true.

What does it take for a moral generic to be true? This is a difficult question. For our purposes, it is sufficient to recognize that some moral principles express generic generalizations,[10] and that some generic generalizations can remain true even if many or most members of the kind lack the property in question, and even if the conditions under which members of the kind lack the property cannot be precisely specified. Since generics admit of exceptions that are not counterexamples, moral principles which express generics are compatible with particular judgments that would be inconsistent with the corresponding universal. Thus, moral principles that express generic generalizations do not generate familiar dilemmas. (At the end of the paper we consider how to treat genuine counterexamples to generics, if and when they arise.)

Section 3: The Generic Overgeneralization Effect

Given that *universal* generalizations are the kinds of generalizations that generate familiar epistemic dilemmas, and given that some of the most popular justifications for taking the generalist route in response to these dilemmas is that our beliefs in certain universal generalizations are non-inferentially justified, one potentially fruitful way to advance the debate between generalists and particularists would be to look more closely at the psychological processes underlying people's evaluation of universal generalizations. Under what conditions do people accept universal generalizations, and under what conditions do they reject them? Recent work in cognitive science suggests that people accept universal generalizations that are false when there are nearby generic generalizations that are true. Leslie, Khemlani, and Glucksberg (2011) call this "the *generic overgeneralization* (GOG) effect, since it involves overgeneralizing from the truth of a generic to the truth of the corresponding universal statement." (p. 17)

The GOG effect was first documented in young children. Hollander, Gelman, and Star (2002) found that 3-year-old children agreed to universal generalizations just as often as they did to corresponding generic generalizations. For instance, when asked, "Do all books have color pictures?" 3-year-olds responded "yes" just as often as they did when they were asked, "Do books have colored pictures?" This GOG effect appears to be due to the 3-year-olds assimilating the universal versions of the questions to the generic versions rather than the other way around; the 3-year-olds' responses to the generically formulated questions were statistically indistinguishable from the responses of 4-year-olds and adults, but their responses to the universally formulated questions differed from the more mature responses of the 4-year-olds and adults. Qualitatively speaking, the responses of the 4-year-olds were intermediate between the responses of the 3-year-olds and the adults to the universal questions, suggesting that the 4-year-olds in this study may also have been susceptible the GOG effect, though to a lesser degree than the younger children. Tardif et al. (2011) replicated these effects among Mandarin-speakers with one difference: Mandarin-speaking children did not begin to discriminate between generics and universals until the age of 5. Similar results have also been found among Quechua-speaking children (Mannheim, Gelman, Escalante, Huayhua, & Puma, 2011).

One might be tempted to think that the reason children treat universal generalizations like generics is that they simply do not understand the meaning of the word “all,” however this is not so. In the studies mentioned above, children were also presented with a set of crayons and asked whether the crayons in front of them were *all* in a box. Across trials, experimenters varied the number of crayons that were in the box as opposed to on the table. In the vast majority of trials, children were able to correctly evaluate whether all of the crayons were in the box. This shows that children do in fact understand the meaning of the word “all.” Despite clearly understanding the meaning of the word “all” when “all” quantifies over a given set of concrete items, children are unable to distinguish universal generalizations about various kinds from generic generalizations about those kinds, taking the universals to be true whenever their corresponding generics were.

Another possibility is that the GOG effect found in these studies is due to 3-year-olds simply being unaware of the counterexamples to these generalizations. Although this is unlikely—3-year-old children presumably know there are books without color pictures—a further study tells against this hypothesis. Leslie and Gelman (2012) presented 4-year-olds and adults with a series of pictures of novel animal kinds. Each picture had eight individuals of the same kind, and the individuals were all identical except that two lacked some target feature that the other six had. For instance, one picture presented eight “gorps,” and all of the gorps looked the same, except that six of the gorps had curly hair and two did not. Participants were asked to demonstrate their knowledge that all of the individuals in the picture were gorps by pointing to each gorp. Once it was clear that participants knew that all of the individuals in the picture were gorps, participants were asked three questions about the relationship of the target feature to the kind: “Do gorps have curly hair?”, “Do all of *these* gorps have curly hair?”, and “Do all gorps have curly hair?” These three questions differed in the nature of the generalization being asked about; the first asks about a generic generalization, the second about a universal that pertains only to a specific set of individuals, and the last about a kind-wide universal.

The results show that the 4-year-olds displayed the GOG effect a substantial portion of the time: 4-year-olds endorsed the kind-wide universals on 51% of trials, while they endorsed restricted universals on only 21% of trials. Despite having counterexamples to the kind-wide universals right in front of them, and despite using those counterexamples to reject the universals that pertained only to a specific set, many 4-year-olds seem to completely disregard those counterexamples when evaluating the corresponding kind-wide universals. A compelling explanation for this surprising result is that 4-year-olds evaluated the kind-wide universals as if they were generics, which can remain true in the face of counterexamples. Because the counterexamples to the kind-wide universals were as salient as they could possibly be, these results provide an unusually powerful demonstration of the GOG effect.

These three studies establish only that young children are susceptible to the GOG effect. Indeed, the adults in these studies decidedly did not display the GOG effect. One possibility is that this is the sort of error we outgrow, but another possibility is that, under different—perhaps more demanding—circumstances adults will also be susceptible to the effect. A theoretical position that has recently gained some traction holds that the processing of generics is cognitively more fundamental than the processing of universal generalizations (e.g., Gelman, 2010; Johnston & Leslie, 2012; Leslie, 2007, 2008, 2012; Leslie & Gelman, 2012; Leslie et al., 2011; Mannheim et al., 2011; Meyer, Gelman, & Stilwell, 2010; Tardif et al., 2011). According to this hypothesis—often termed the "generics-as-defaults" hypothesis—generic generalizations reflect the default mode of generalizing in humans, whereas quantified statements such as universal generalizations reflect a non-default mode of generalizing. While the ability to form generic generalizations is early-developing, automatic, and "primitive," the ability to form quantified generalizations is later-developing, cognitively more demanding and sophisticated, and furthermore may require that the disposition to form a generic generalization instead must be inhibited (Leslie, 2007, 2008, 2012; Leslie et al., 2011). This hypothesis naturally explains the findings of Hollander et al., Mannheim et al., and Tardif et al.; even though preschool children have a basic *competence* with universal quantifiers, when they are asked to consider kind-wide universal generalizations, they may instead rely on their understanding of the less taxing generic generalization, and so in effect substitute their evaluation of a generic for their evaluation of a universal. Since these sorts of tendencies are rarely fully outgrown, one would also expect that adults, like young children, would at times fail to inhibit this tendency to 'fall back' on generic generalizations.

A number of studies confirm that there are circumstances under which adults exhibit the GOG effect. Leslie, Khemlani, and Glucksberg (2011) asked adult participants to evaluate the truth of a number of generalizations such as "Triangles have three sides," "Tigers have stripes," "Lions have manes," "Cars have radios," "Pit bulls maul children," and "Canadians are right-handed." Each generalization appeared either as a universal generalization (e.g., "All ducks lay eggs"), a generic generalization (e.g., "Ducks lay eggs"), or an existential generalization (e.g., "Some ducks lay eggs.") The question of interest is whether participants would accept universal generalizations that were false when the corresponding generic generalizations would have been true.

To exhibit the GOG effect, participants would have had to accept universal generalizations that were *false* when these universals had corresponding generic generalizations that were *true*. For a certain class of generalizations, this is just what was found. When the generalizations ascribed a *characteristic* property to members of the kind—"a property that bears a deep causal and explanatory relation to the kind in question"—participants were as likely to accept the universal generalization as they were to reject it (Leslie et al., 2011, Experiment 1). When the characteristic property was a property possessed by a majority of the kind (e.g., tigers having stripes), then participants accepted the universal generalization 78% of the time. When the characteristic property was a property possessed by a minority of the kind (e.g., lions having manes), then participants accepted the universal generalization 51% of the time. Moreover, participants consistently made the mistake, and the mistakes were not due to just one or two problematic items: 40 out of 56 participants made the mistake over 30% of the time with 'minority characteristic' generalizations, and mistakes were made over 30% of the time with 10 out of the 12 minority characteristic generalizations. These results strongly suggest that people are prone to committing the GOG error when the properties ascribed in the universal generalizations are characteristic of the kind in question (Leslie et al., 2011, p. 21).

Leslie et al. carried out a number of follow-up experiments to rule out various alternative explanations for the effect. One such explanation holds that people endorse the false universal generalizations not because they are committing the GOG error, but because they take the universal to be quantifying over subkinds of the kind mentioned in the generalization (i.e., to be asserting that all *subspecies* of lions (Asiatic lions, West African lions, etc.) satisfy the generic "Xs have manes"), or to involve a contextual restriction to a subset of lions (e.g., to male lions). In order to rule out these hypotheses, Leslie et al. replicated their first experiment, except that they provided participants with information about the size of the population of the kind mentioned in the generalization before participants evaluated the generalization. For instance, before evaluating the universal generalization "All ducks lay eggs," participants were told: "Suppose the following is true: there are 431 million ducks in the world." This change was intended to ensure that participants understood the universal to range over each and every individual member of the kind. The GOG effect persisted even under these circumstances, suggesting that it is not simply a matter of subkind quantification or domain restriction (Leslie et al., 2011, Experiment 2a).[11]

Another alternative explanation for the GOG effect is that people endorse false universal generalizations not because they are committing the GOG error, but because they simply do not realize that these generalizations are subject to counterexamples. To test this, Leslie et al. had participants evaluate generalizations just as before, but also had them take a "knowledge test." The knowledge test included items that would allow participants to display whether they knew that the minority characteristic generalizations were subject to counterexamples, since only one sex of each kind had the relevant property. For instance, items on the knowledge test included "Male ducks lay eggs" and "Female lions have manes." If participants rejected these claims but continued to exhibit the GOG effect, this would rule out the hypothesis that participants accepted the generalizations because they were ignorant of its counterexamples (Leslie et al., 2011, Experiment 3).

Half of all participants took the knowledge test before evaluating the generalizations while the other half evaluated the generalizations before taking the knowledge test. When participants evaluated the generalizations first, participants accepted the relevant universal and then went on to judge that only one sex has the property on 40% of trials.[12] These responses cannot be dismissed as due to ignorance of counterexamples, because participants' subsequent responses on the knowledge test revealed that they were generally aware of the fact that these generalizations were subject to counterexamples. When participants took the knowledge test first, however, they accepted the relevant universal despite having just correctly judged that one sex lacks the property on 19% of trials. Thus, reminding participants of counterexamples before asking them to evaluate the generalizations reduced the rate at which they committed the GOG error, but it did not eliminate it. (Leslie et al., 2011, Experiment 3)

Interestingly, a very different effect was found with 'majority characteristic' universals—universals that ascribe a characteristic property to every member of a kind when most (but not all) members of that kind possess that property (e.g., "All tigers are striped"). When participants evaluated the generalizations first, they committed the GOG error with these majority characteristic universals on 70% of trials, but when they took the knowledge test first, they committed the GOG error with majority characteristic universals on 90% of trials. Leslie et al. speculate that "Majority characteristic universals may have been accepted more often to potentially compensate for the reduction in agreement to minority characteristic universals [e.g., "All lions have manes"]." (2011, p. 25) Nevertheless, this is a striking result, as one might have predicted that alerting participants to the fact that there were exceptions to minority characteristic universals might have indirectly alerted them to the fact that there were exceptions to the majority characteristic universals, decreasing the rate at which they accepted majority characteristic universals.[13]

The results of these experiments suggest that adults are susceptible to the GOG effect when evaluating characteristic universals, and that they remain susceptible even when they recognize that the generalizations are subject to counterexamples. They also suggest that adults are especially susceptible to the GOG effect when evaluating majority characteristic universals such as "All dogs have four legs"—universals that ascribe a property that is characteristic of the kind to every member of the kind, when in fact a majority—but not all—members of the kind possess that property.

Other studies provide additional support for these conclusions. Meyer, Gelman, and Stilwell (2010) asked participants to evaluate a series of universal and generic generalizations, but half of their participants were told to answer as quickly as possible. When told they could take as long as they would like, participants were able to evaluate false majority characteristic universals such as "All dogs have four legs" with a reasonable level of accuracy. When told to answer as quickly as possible, however, participants' mean level of accuracy in evaluating false majority characteristic universals dropped off substantially (Meyer et al., 2010, p. 916). In other words, when speeded, participants evaluated false majority characteristic universals as if they were generics (i.e., as true) more often than they did when they had more time. Importantly, no corresponding effect was found on participants' accuracy in evaluating majority characteristic *generics*; participants continued to evaluate majority characteristic generics with a fairly high level of accuracy even when they were rushing (p. 916). This suggests that adults are more likely to fall victim to the GOG effect when they are under time pressure, and that this is due to a special feature of these universals rather than a general inability to evaluate generalizations accurately under time pressure.

Meyer et al. also found that participants under time pressure took longer to correctly reject false majority characteristic universals (e.g., "All dogs have four legs") than they did to correctly accept true majority characteristic generics (e.g., "Dogs have four legs"). Furthermore, they found that participants under time pressure took longer to correctly reject false majority characteristic universals than to correctly reject "irrelevant-scope" generalizations that, unlike majority characteristic universals, do not have true corresponding generics (e.g., "All squirrels have beaks") (2011, p. 917). These results suggest that rejecting a false majority characteristic universal like "All dogs have four legs" takes longer than accepting its true generic counterpart not just because rejecting a false generalization generally takes longer than accepting a true generalization, but because rejecting a false *majority characteristic* universal may require overcoming an initial inclination to evaluate the universal as if it were a true generic. This hypothesis is further supported by the finding that, regardless of whether they were under time pressure, participants took longer to accurately accept true majority characteristic universals than to accurately accept true majority characteristic generics (p. 916). Meyer et al. speculate that even when unrushed, "participants were likely hesitant to endorse the veracity of the universally-quantified sentences because they were trying to think of plausible counterexamples (although they eventually did tend to respond *true*, as predicted), whereas they were relatively fast to respond to the generic version based on their default generic representations." (p. 917)

These studies suggest the following picture of what happens when people evaluate the truth of a universal generalization. First, they draw upon their representation of the corresponding generic generalization. If the corresponding generic is true, then they feel an inclination to accept the universal. If time allows, they then search their semantic memory for counterexamples. If after a contextually determined amount of time they fail to find a counterexample, they accept the universal; if they find a counterexample within the allotted time, they reject it.

This proposal involves two central claims: (1) the initial intuitive attraction to universal generalizations is driven by endorsement of the corresponding generics, and (2) whether people go on to endorse the universal generalization depends on whether they can find a counterexample. These claims enjoy independent support from a number of studies on adult reasoning and memory.

Take the claim that people's initial attraction to universal generalizations is driven by their endorsement of the corresponding generic. If this claim is to find support, there must be cases in which people exhibit a pattern of belief that they would only exhibit if they were assimilating universal generalizations to their corresponding generics. There are such cases.

Consider what has come to be called the "inverse conjunction fallacy," which is the error of failing to attribute some property to all members of a conjunctively defined subset of a larger group when one nevertheless believes that all members of the larger group have that property (Jönsson & Hampton, 2006). For instance, imagine one accepts the following claim: "All lambs are friendly." (p. 332) One would be committing the inverse conjunction fallacy if one were unwilling to also accept "All dirty German lambs are friendly." (p. 332) Since dirty German lambs are merely a conjunctively defined subset of lambs, one should be willing to attribute any property to every member of the subset that one is willing to attribute to every member of the superset. Jönsson and Hampton (2006) found that people are not willing to do this. They were far less willing to accept sentences like "All dirty German lambs are friendly" than they were to accept sentences like "All lambs are friendly."

As Jönsson and Hampton suggest, we can make sense of the inverse conjunction fallacy if we assume that people are evaluating the universal generalizations as if they were generics. Because generics tolerate exceptions, there is nothing inconsistent about endorsing "Lambs are friendly" but rejecting "Dirty German lambs are friendly." "Lambs are friendly" can remain true even if there are lambs that are not friendly, and it may well be that the class of lambs that dirty German lambs tend to fall into is the class of lambs that are not friendly. Because people display this same pattern of responses when the generalizations are *universal* rather than generic, this suggests that they are implicitly evaluating the universals as if they were generics. Jönsson and Hampton conclude that "the addition of universal quantifiers to the statements is likely to change people's judgments very little. [. . .] Universally quantified sentences and generic sentences are likely to be treated similarly unless the context very clearly supports extensional reading." (2006, p. 319)

Likewise, Steven Sloman (1993, 1998) found that people have a tendency to find arguments like

> All plants contain bryophytes, therefore all flowers contain bryophytes.

to be more convincing than arguments like

> All plants contain bryophytes, therefore all mosses contain bryophytes.

This pattern of response is puzzling because both flowers and mosses are plants. If both flowers and mosses are plants, then the premise that all plants contain bryophytes supports the claim that all flowers contain bryophytes just as well as it supports the claim that all mosses contain bryophytes. However, this pattern of response becomes comprehensible if we suppose that participants are interpreting the universal generalizations as generics. "Plants contain bryophytes" may support the claim that flowers contain bryophytes more strongly than the claim that mosses contain bryophytes if one has more reason to believe that moss may be an exception to the rule than flowers are (e.g., perhaps one thinks that flowers tend to have typical plant-properties, while mosses are somewhat odd and unpredictable). (Leslie, 2012)

Together with Meyer, Gelman, and Stilwell (2010) and Leslie, Khemlani and Glucksberg (2011), these studies support the claim that people's attraction to universals is often due to their endorsement of generics. Of course, people do not always go on to *endorse* the universal; the results from Meyer et al. (2010) and Leslie et al. (2011) suggest that people do not typically endorse the universal if they discover a counterexample to it. This in turn suggests that people may refrain from endorsing universals that have true corresponding generics only if they can find a counterexample to the universal. While the results from Meyer et al. and Leslie et al. do not settle the question, this claim enjoys independent support from research on conditional reasoning.

A number of studies show that people's evaluation of a conditional argument depends not only on the form of the argument, but on how plausible they find its conditional premise.[14] For this reason, studies of this sort can be used to investigate how the retrieval of counterexamples influences people's willingness to accept conditionals. One such study comes from De Neys, Schaeken, and d'Ydewalle (2003). De Neys et al. found that participants were less willing to use modus ponens to draw a conclusion from a conditional such as, "If the ignition key is turned, then the car starts," as the number of counterexamples (e.g., "The ignition key is turned, but the car doesn't start because it's out of gas") they considered increased. In one experiment, these counterexamples were explicitly presented alongside the conditionals; in a second experiment, these counterexamples were produced by the participants themselves a month before they were asked to report their willingness to use the conditional to draw a conclusion via modus ponens. In both cases, the more counterexamples people considered, the less confident they were that they could draw the conclusion that the consequent of the conditional was true from the assumption that its antecedent was true.

Assuming participants did not come to doubt the validity of modus ponens as a rule of inference, these results suggest that the retrieval of counterexamples reduced people's willingness to endorse the conditionals in question. Since these conditionals (e.g., "If the ignition key is turned, then the car starts") were equivalent to or instantiations of majority characteristic universals (e.g., "All acts of turning the ignition key result in the car starting"; Kratzer, 1991) with true generic counterparts (e.g., "Acts of turning the ignition key result in the car starting"), these results provide further evidence that the retrieval of counterexamples plays a key role in motivating people to reject universals they would have otherwise accepted on the basis of accepting the corresponding generics.[15]

To defend the view that people endorse universals that have true corresponding generics when they fail to retrieve counterexamples to the universal, we would have to show that counterexample retrieval is not an entirely automatic process. We would have to show that counterexample retrieval is a cognitively demanding process that must compete with and overcome people's initial inclination to accept false universals.

The fact that exposing people to counterexamples reduced the GOG effect in Leslie et al. (2011) provides initial evidence for this conclusion, as does Meyer et al. (2010)'s finding that putting people under time pressure increases the GOG effect.[16] Further evidence for this claim comes from studies showing that counterexample retrieval depends largely on working memory. For instance, De Neys, Schaeken and d'Ydewalle (2005a) found evidence that the process of retrieving counterexamples depends heavily on working memory capacity (see also DeNeys, Schaken & d'Ydewalle, 2005b for additional evidence). They found that people who were higher in working memory capacity were able to retrieve more counterexamples to conditionals in a limited amount of time than those who were lower in working memory capacity. They also found that burdening people's working memory with a secondary task—continuously tapping out an unusual pattern with their fingers—significantly decreased the number of counterexamples they could retrieve in the allotted amount of time. These findings lend support to the claim that counterexample retrieval is a cognitively demanding process that must compete with and overcome people's initial inclination to accept false universals.[17]

We have argued in this section that people are often susceptible to the GOG effect, and that the GOG effect is a byproduct of the way people typically evaluate universal generalizations. In particular, we have argued that people typically evaluate universal generalizations by first seeing whether the universals have true corresponding generics and then, if so, searching for counterexamples. When universals do have true corresponding generics, people find the universals initially attractive. When people find counterexamples to the universals, they dismiss the universals despite their initial attractiveness. However, if people cannot find counterexamples to the universals or have some independent reason for disregarding any apparent counterexamples, they go on to endorse the universals on the basis of their original attraction.[18] This attraction is due to the fact that these universals have true corresponding generics—not due to the fact that people consult discrete, set-theoretic representations that vindicate the universal.

Section 4: The Moral GOG Effect

People often display the GOG effect when evaluating universal generalizations like "All ducks lay eggs" and "All pencils are wooden." In this section, we argue that we should expect them to do the same when evaluating universal generalizations like "Lying is always wrong" and "Killing is always impermissible."

One point to note at the outset is that, if the generics-as-defaults hypothesis is correct—i.e., if indeed our basic way of forming generalizations is generic rather than quantificational in nature—then this means that the moral principles we acquire as children will almost certainly be generic rather than universal in nature. Further, there is no evidence to suggest that these principles will be overwritten with universals when we become adults—there is scant evidence in psychology that we *ever* represent such core generalizations as universals, outside perhaps of special cases such as mathematics (see Johnston & Leslie, 2012, for a summary of findings). The balance of empirical evidence would suggest that the moral principles that would naturally spring to mind—and strike us as self-evident, as platitudes, or simply as intrinsically plausible—would surely be generic in nature. Might we fall into thinking that the corresponding universals are true, even if they are not?

We saw in the last section that people are prone to committing the GOG error when evaluating *characteristic* universal generalizations, such as "All ducks lay eggs" or "All tigers are striped." The question that naturally arises, then, is whether moral generalizations are characteristic generalizations. A characteristic property will be—if the notion is to be of any interest or value—a hypothesized natural type of property that cognitive psychology needs to recognize. In this way, a full characterization of the notion of a characteristic property is best given in light of considerable empirical investigation, rather than stipulated beforehand. However, for our purposes, we may gloss a characteristic property as being one that is believed to have a deep, causal or otherwise explanatory connection to the kind in question—one that is believed to result from the inherent nature the kind. (In this way characteristic properties contrast with more accidental ones—properties that we believe individuals to possess only as a result of extrinsic, more adventitious circumstances or interventions. Examples of generics involving such accidental or adventitious properties include "Barns are red" and "Cars have radios.")

It seems clear that moral generalizations are naturally understood as involving characteristic properties in this sense. While the idea that a kind's moral properties are somehow caused by the nature of the kind is almost universally rejected, it is nearly universally held that the moral properties of a kind are nevertheless *explained by* or and *a result of* the non-moral nature of that kind. At least, this is uncontroversially true of the kinds of moral generalizations that concern ethicists. Practitioners of the method of reflective equilibrium, for instance, are not content to identify a set of moral principles that merely imply their considered judgments; they seek principles that also *explain* them.[19]

Work by Sandeep Prasada and colleagues has produced several tests that can be used to indicate whether a property is taken as characteristic a kind (Prasada & Dillingham, 2006, 2009; Prasada, Khemlani, Leslie, & Glucksberg, 2013; passing (most of) these tests should be understood as indicators of a property's being characteristic of a kind, not constitutive of the fact). In particular, Prasada and colleagues' work suggests that people are significantly more likely to agree to certain statements when the property in question is characteristic of the kind. As an illustration, consider the following trios:

1a) Tigers, by virtue of being tigers, have stripes.
b) Lions, by virtue of being lions, have manes.
c) Cars, by virtue of being cars, have radios.

2a) Having stripes is one aspect of being a tiger.
b) Having a mane is one aspect of being a lion.
c) Having a radio is one aspect of being a car.

3a) Tigers are supposed to have stripes.
b) Lions are supposed to have manes.
c) Cars are supposed to have radios.

Participants robustly accept the a- and b-sentences as true to a greater extent than they do the c-sentences, across a range of items (Prasada et al., 2013). Participants also judge the first two explanations to be significantly better than the third (Prasada & Dillingham, 2006, 2009; Prasada et al., 2013):

Q: Why does that [pointing to a tiger] have stripes?
A: Because it is a tiger.

Q: Why does that [pointing to a lion] have a mane?
A: Because it is a lion.

Q: Why does that [pointing to a car] have a radio?
A: Because it is a car.

These findings suggest that these tests may be helpful in deciding whether moral generalizations should be understood as characteristic. While a full investigation of this would involve running Prasada et al.'s (2013) experiments on a full set of items, consideration of a few examples suggests that moral generalizations—here reworked to best parallel Prasada et al.'s items—should be classified as characteristic. In the trios below, the first two examples—the moral examples—would seem to fare considerably better in these formulations than the third example (which, while seemingly not *characteristic* of the acts in question, nonetheless involves an all too prevalent property of them!).

1d) Acts of lying, by virtue of being acts of lying, are wrong.
e) Acts of killing innocents, by virtue of being acts of killing innocents, are wrong.
f) Acts of grading papers, by virtue of being acts of grading papers, are boring.

2d) Being wrong is one aspect of being an act of lying.
e) Being wrong is one aspect of being an act of killing innocents.
f) Being boring is one aspect of being an act of grading papers.

3d) Acts of lying are supposed to be wrong.
e) Acts of killing innocents are supposed to be wrong.
f) Acts of grading papers are supposed to be boring.

Q: Why is that [indicating an act of lying] wrong?
A: Because it is an act of lying.

Q: Why is that [indicating an act of killing innocents] wrong?
A: Because it is an act of killing innocents.

Q: Why is that [indicating an act of grading papers] boring?
A: Because it is an act of grading papers.

It would seem reasonable to conclude that these moral generalizations are characteristic generalizations.

Another feature of characteristic generalizations is that people are often willing to arrive at them on the basis of very limited evidence—in contrast to more 'accidental' generalizations, for which more evidence is required (Leslie, 2008; Nisbett, Krantz, Jepson, & Kunda, 1983). In a review of the literature, Darley and Schultz (1990) find that this is the most common means by which children learn moral rules:

> Children are disposed to extract rules from the observation of concrete instances. In the domain of moral reasoning they are abetted in this endeavor when adults explicitly extract the general rule for the child. ("Well then, since you did it by accident, I won't punish you.") Observational evidence suggests that adults perform this service infrequently; children must often infer the general rule that lies behind their seniors' specific prohibitions. (p. 544)

The fact that even young children are able to extract moral rules from observation of concrete instances suggests that they take the moral properties of the concrete instances to be characteristic of the abstract kind of which they are instantiations. (For a review of developmental evidence concerning the emergence and acquisition of moral thinking in young children and infants, see Bloom (2013).)

Further evidence that moral generalizations pattern in this way comes from a clever study on adults by Jay Van Bavel and colleagues. Van Bavel et al. (2012) asked participants to evaluate many different kinds of actions, such as adopting a child, eating fast-food, and carrying a concealed knife. In some cases, participants were asked to judge whether it would be *morally* good or bad for them to perform the action themselves. In other cases, participants were asked to judge whether it would be *non-morally* good or bad (i.e., pragmatically or hedonically good or bad) for them to perform the action themselves. After evaluating their own imagined performance of the action, participants were then asked whether this action should be "universally prohibited" or "universally required," where being universally prohibited means "that nobody should be permitted to do this action, without exception" and where being universally required means that "everybody should be required to do this action, without exception." (Van Bavel et al., 2012, p. 8)

When participants were asked to morally evaluate their own imagined performance of the action, their moral evaluations were more highly correlated with their universality judgments than when they were asked to non-morally evaluate their imagined performance of the action. In other words, if you ask people whether it would be non-morally good or bad for them to perform an action, and they believe it would be non-morally good to a given degree, they will be less likely to believe that action is universally required than people who are asked whether it would be *morally* good or bad for them to be perform that action and who believe it would be *morally* good to the same degree. Likewise, if you ask people whether it would be non-morally good or bad for them to perform an action and they believe the action is non-morally bad to a given degree, they will be less likely to believe the action is universally prohibited than people who are asked whether it would be *morally* good or bad for them to perform that action and who believe it would be *morally* bad to the same degree.

In addition to finding this effect of construal (moral vs. non-moral) on universality judgments, Van Bavel et al. also found an effect of construal on the time it took participants to make their universality judgments. Participants who were asked whether it would be morally good or bad for them to perform the action made universality judgments *faster* than those who were asked whether it would be non-morally good or bad for them to perform the actions.

Although Van Bavel et al. initially found these effects *across* participants, they found that the effects remain even when the very same people evaluate the very same action both morally and non-morally. These results suggest that people take moral properties to be even more characteristic of a given kind than other evaluative properties, for people are more willing to endorse universal generalizations regarding the obligatoriness or impermissibility of a particular kind of act—and do so more quickly—after considering the moral properties of one such act than after considering other evaluative properties of that act. The moral evaluation of a particular action seems to more easily generalize than other types of evaluation. These findings, along with Prasada and colleagues' tests, suggest that people take the moral properties of a particular kind of action to be characteristic of that kind.

Section 5: The Argument Against Generalism

The studies canvassed in section 3 show that people are susceptible to the GOG error when evaluating characteristic universals. In section 4, we argued that the moral principles at issue in moral theorizing are clearly intended to ascribe *characteristic* moral properties to various morally-evaluable kinds. It follows, then, that we should expect people to commit the GOG error when evaluating moral universals: That is, we should expect for people to regularly believe universal moral generalizations are true even when they are false, simply because they find them intrinsically plausible and so difficult to reject, thanks to the plausibility of the related generic.

The method of considering the corresponding generic and then searching for counterexamples, need not in itself be unreliable. However, the reliability of the method depends on the *extent* of the search for counterexamples. Where the search for counterexamples is abruptly terminated by time pressure or other cognitively demanding tasks, the method is unreliable. Moreover, where the search for counterexamples is terminated *simply because* one finds the generalization itself to be intrinsically plausible then that method is unreliable, for it simply reflects the plausibility of the corresponding generic, as in the GOG effect. Unfortunately, this truncation of the reliable form of the method seems typical of generalism, and is manifested in generalists' use of moral universals to override intuitions about particular cases.

Reflection on the GOG effect and like phenomena suggests that, when we believe a universal moral generalization is true simply because we find it intrinsically plausible, we should recognize that this belief was formed by a process that regularly leads to false beliefs. It is a plausible and widely held epistemic principle that one cannot be justified in holding a belief when one is justified in believing that the belief was formed by a process that regularly leads to false beliefs, and one has no other independent grounds that would justify the belief.[20] Assuming this principle is correct,[21] certain justifications for epistemological generalism are undermined. In particular, it follows that we will not be justified in believing that universal moral generalizations are true, if we believe them to be true simply because we find them intrinsically plausible or difficult to reject.[22]

Of the three justifications for generalism discussed in section 1, seeming state theory is the one most clearly undermined by this argument. Recall that according to seeming state theory, we are justified in endorsing a general moral principle just in case that principle seems to be true and we lack any defeaters for that belief. We do in fact have a defeater for that kind of belief: We can provide a plausible explanation for why moral universals seem to be intrinsically plausible without assuming at any point that they are true. This explanation is sufficiently probable to undermine the justification we thought we had for endorsing these moral universals on this basis.

This argument also undermines one of the justifications for generalism offered by those who are practitioners of the method of reflective equilibrium. According to this sort of generalist, we are justified in endorsing a general moral principle when we have more confidence in the principle than we do in any of our incompatible judgments about particular cases. To the extent that this confidence is simply due to finding the principle intrinsically plausible or difficult to reject, the argument above robs us of our justification for favoring the moral principle over the particular judgment.

Finally, consider self-evidence theory. According to self-evidence theorists, we are justified in endorsing a self-evident moral principle when and because we have an adequate understanding of that principle:

> It seems to me self-evident that, other things being equal, it is wrong to take pleasure in another's pain, to taunt and threaten the vulnerable, to prosecute and punish those known to be innocent, and to sell another's secrets solely for personal gain. When I say such things, I mean that once one really understands these principles (including the ceteris paribus clause), one doesn't need to infer them from one's other beliefs in order to be justified in thinking them true. (Shafer-Landau, 2003, p. 248)[23]

Like seeming state theorists, contemporary self-evidence theorists concede that self-evident beliefs are defeasible. But do self-evidence theorists arrive at their beliefs about moral principles via a psychological process whose reliability is undermined by the argument offered here? Do self-evidence theorists endorse the moral principles they do on the basis of their intrinsic plausibility? Although self-evidence theorists often talk in terms of intrinsic plausibility and its ilk, they typically refrain from *defining* their position in those terms. For instance, here is Shafer-Landau discussing the putatively self-evident principle that, "absent special circumstances, one's enjoyment doesn't make it right to hurt others":

> There certainly seems to be something intuitively plausible about such a claim. In fact, despite much talk of the revocability of all of our judgments, this principle seems so plausible as to resist any efforts at revision. I would be more inclined to view a system that rejected it as corrupt, as a counterfeit morality, than to abandon such a claim. This resistance to doxastic alteration is itself neither strictly necessary nor sufficient for a belief to qualify as self-evident. But something like this degree of resolve is what we would expect to see attached to any self-evident belief. (2003, p. 249)

Even if self-evident beliefs need not enjoy intrinsic plausibility, resistance to revision, or any other introspectible quality to count as self-evident, it seems that people typically do endorse putatively self-evident moral principles on the basis of their having some such quality (in addition to or as a consequence of their having adequately understood the believed proposition). To the extent that this is true, we are justified in believing that these beliefs are the products of an unreliable process. And as long as we have no way of distinguishing genuinely self-evident beliefs that are the output of this process from beliefs that merely appear to be self-evident, our non-inferential justification for endorsing them is undermined. Therefore, the self-evidence theorists' justification for endorsing general moral principles is also threatened by the argument presented in this paper.

A self-evidence theorist may object, however, that their endorsement of general moral principles has nothing to do with their having found those principles to be intrinsically plausible. According to this kind of self-evidence theorist, their endorsement of a general moral principle has to do solely and directly with their having adequately understood the principle. Although we find this claim psychologically implausible, we will argue that the studies from section 3 still pose a challenge to this kind of self-evidence theorist.

Recall that, in section 3, we argued that when people endorse a characteristic universal, they do so *not* because they explicitly process the set-theoretic relations that constitute the content or meaning of the universal, but because the characteristic universal has a true corresponding generic. Even if people do adequately understand characteristic universals, their endorsement of the universals does not *depend* on their adequate understanding of these universals. It depends on the fact that the universal's corresponding generic is true. While people *can* come to reject characteristic universals with true generics, they typically do so because they bring to mind some particular counterexample to the universal—not because further reflection on the universal itself or a more adequate understanding of the meaning of the universal reveals it to be false.

Of course these claims depend in large part on what 'adequate understanding' amounts to. If adequately understanding a principle can essentially involve searching for counterexamples, then self-evidence theorists can claim to have adequate understanding of general moral principles. Jeff McMahan, for instance, accepts the view that adequately understanding a principle requires knowing what it entails:

> To be justified in accepting a moral principle, we must first understand what it commits us to in particular cases. As William James noted in a letter written long before he became a practicing philosopher, "No one sees farther into a generalization than his own knowledge of the details extends" (Barzun, 1983, p. 14). So, while I regard the principle rather than our intuitions as foundational, I do not think that moral inquiry can proceed by deducing conclusions about particular cases from self-evident moral principles. Rather, *the order of discovery is the reverse of the order of justification.* (2013, p. 114)

While McMahan's view can help the self-evidence theorist explain how we can come to have adequate understanding of moral principles, it does so at the cost of making self-evidence theory unavailable as a justification for epistemological generalism. Recall that epistemological generalism—as it is understood in this paper—is the view that we typically have reason to reject our judgments about particular cases when they conflict with our judgments about general principles. If adequately understanding a general principle requires taking our intuitions about particular cases as data concerning the truth of that general principle, then it will never be the case that we both accept a general principle on the basis of adequately understanding it and hold an incompatible judgment about a particular case; our adequate understanding of the principle is in part *manifested* by our judgment about the case. It follows from a view like McMahan's that self-evidence theory cannot support generalism about moral epistemology. Familiar epistemic dilemmas are not as they seem to the generalist. Instead they typically represent positive results in the process of counterexample search. Overgeneralizations have thereby been exposed as such.

Section 6: Objections and Replies

Objection: Even granting that adequate understanding of a moral principle requires consideration of apparent counterexamples, it does not follow that recognition of these apparent counterexamples always defeats our justification for endorsing the principle. If the principle is self-evident, our justification for endorsing the principle derives from the fact that we adequately understand it. If we've considered all of the apparent counterexamples to a self-evident principle, then we've adequately understood the principle, and if we continue to find the principle attractive after adequately understanding it, we have justification for endorsing it.

Reply: Even granting that consideration of all the apparent counterexamples to a principle is sufficient for adequately understanding that principle, it does not follow on most self-evidence theories that one can be justified simply in virtue of understanding the principle. This is because self-evidence theories require that your belief in a moral universal be *based* on—that is, *caused* by—an adequate understanding of the universal, and the psychological evidence we've presented suggests that this is rarely the case. The evidence we've offered suggests that people's attraction to moral universals is based on their initial tendency to equate the universal with its generic counterpart, *not* on their recognition of apparent counterexamples. If anything, people's recognition of apparent counterexamples *decreases* their attraction to universal generalizations. This suggests that considering apparent counterexamples plays no role in producing one's endorsement of the universal. Even if the universal remains attractive in the face of apparent counterexamples, that lingering attraction seems to be due to one's initial treatment of the universal as a generic rather than one's later, adequate understanding of the universal as having counterintuitive consequences for particular cases.

Objection: Unlike the kinds of generalizations that have been shown to lead to the GOG error in psychological studies, people do not reject *moral* universals as soon as they learn of an apparent counterexample. While people will immediately concede that "All tigers have stripes" is false as soon as they see a stripeless tiger, people do not immediately concede that "Lying is always wrong" is false once it is pointed out to them that lying does not seem to be wrong in certain circumstances. This suggests that people's attraction to moral universals is not a matter of overgeneralization.

Reply: According to this objection, attraction to a universal generalization can be the product of overgeneralization only if one immediately rejects that universal upon recognition of an apparent counterexample. But this is a mistake. Contrary to what this hypothesis would predict, there are circumstances under which people fail to reject false universals even when they have been reminded of the fact that there are counterexamples to these generalizations (Jönsson & Hampton, 2006; Khemlani & Johnson-Laird, 2012, Experiment 2; Leslie et al., 2011, Experiments 3) For example, as Huemer has noted, “most people who consider the comprehension axiom of naive set theory [a false characteristic universal with a true corresponding generic] find it intuitive (it *seems* right), even those who know the axiom to be false because of the paradoxes it engenders.” (2008, p. 371) These considerations suggest that GOG-driven attraction to false universal generalizations often remains even after people recognize the existence of apparent counterexamples; whether people take this lingering attraction to the universal as evidence of its truth is orthogonal to whether the attraction itself is due to overgeneralization. Thus, the fact that people do not immediately reject moral universals when confronted with apparent counterexamples is no evidence that people’s endorsement of these universals is not ultimately due to overgeneralization.

We claim that people’s endorsement of moral universals is in fact due to overgeneralization, and that the only reason that generalists take their attraction to the universal as evidence of its truth in the face of apparent counterexamples is because of a prior commitment to epistemological generalism. But we have argued that epistemological generalism is untenable. If we are right, epistemological generalism precludes us from the primary way of overcoming the GOG effect—consulting apparent counterexamples. The very methodology of generalism has the peculiar feature of cutting off this avenue of correction. Indeed, when we come up with a putative counterexample to a putative moral universal, the generalist urges us, in the relevant cases, to privilege the (supposedly) universal claim over the counterexample. One might think that this effectively cuts off our best protection against the GOG error.

Objection: Morality is special. Even if people often endorse false characteristic universals because they commit the GOG error, there is special reason to think that they are not committing the GOG error when they endorse moral universals.

Reply: One way morality might be special is that, unlike other domains in which the GOG effect has been found, morality is often thought to be a primarily a priori enterprise. This is, of course, a contested claim. But even if we grant it, it is hard to see why it should make us think that people are less likely to commit the GOG error when evaluating moral universals than other types of characteristic universals. The GOG effect seems to be due to the way the brain represents information concerning kinds and their properties; whether morality is an a priori or a posteriori enterprise seems to be irrelevant. (A further question would also arise here, concerning the nature of this putative a priori knowledge. A familiar way of answering this question has it that the project, like many other philosophical projects, involves the analysis of our concepts. Johnston and Leslie (2012) argue at that length that here again, the centrality of generic, rather than universal, generalizations in our psychology raises difficulties for the enterprise of analysis, as traditionally conceived. Whatever principles or platitudes such a project might uncover will be generic in nature—though of course we may be inclined to perceive them as universals.)

Another way in which morality might be special is that we might have special reason to think that the moral domain is rule-governed in a way that other domains are not. Again, this is a hotly contested claim (see, e.g., Dancy, 2004). Nevertheless, we can grant it for the sake of argument. Does this provide reason to think we might be less likely to commit the GOG error when evaluating moral universals than when evaluating other types of characteristic universals? No. It does not follow from the putative fact that morality is rule-governed that we are any less likely to evaluate a moral universal as true just because its generic counterpart is true. This metaphysical claim about the structure of morality is neutral on the question at hand. If we have good reason to think we are prone to committing the GOG error when evaluating characteristic universals, then we have good reason to believe that we are prone to committing the GOG error when evaluating characteristic *moral* universals—even if we have independent reason to believe that any particular moral fact would have to be grounded in some universal generalization or other.

Objection: Philosophers are special. Even if non-philosophers often endorse characteristic universals because they commit the GOG error, there is special reason to think that philosophers are less susceptible to the GOG error.

Reply: Here's one way in which philosophers might be special. Unlike non-philosophers, competent philosophers are well-trained in the logic of quantified statements. It might seem to follow that they would be less likely to evaluate universal generalizations as if they were generics.

This version of the objection misunderstands the nature of the GOG effect. The GOG effect does not seem to be due to people *explicitly* misunderstanding the truth conditions of universal generalizations; it is due to the way the brain represents kinds and their properties. The key idea is this: When evaluating the truth of universal generalizations, our brains do not typically draw upon discrete, set-theoretic representations of kinds and their properties. Instead, our brains consult representations of a generic form. This should be so regardless of whether people understand the truth conditions for universal statements. At best, understanding the truth conditions for universal generalizations will simply alert people to the need to undertake a thorough search for counterexamples. In any case, recall that Hollander, Gelman, and Star (2002), Tardif et al. (2011), and Leslie and Gelman (2012) found that even very young children who fall victim to the GOG effect are generally competent with the word "all," and adults who display the GOG effect understand in the abstract what it takes for a universal generalization to be true. Having an understanding of quantified statements does not make one immune to the GOG effect.

Another way in which philosophers might be special is that they are smart. And smart people are less prone to cognitive biases.

Or so the story goes. This intuitive thought enjoys only modest empirical support. In fact, recent empirical work on the question actively undermines it. Stanovich and West (2008) found that people who were higher in cognitive ability were no less prone to a wide variety of classic cognitive biases: myside bias, the sunk-cost effect, framing effects, omission bias, affect biases, "less is more" effects, base-rate neglect, outcome bias, the conjunction effect, and certainty effects that are irrational by the standards of expected utility theory. More strikingly, West, Meserve, and Stanovich (2012) found that people who were more intelligent were, if anything, more prone to what has come to be known as the "bias blind-spot": They were more likely to think they were less susceptible to various cognitive biases than others *despite being just as prone to the biases as other people*.

Of course, there are also many biases that intelligent people *are* less susceptible to. While Stanovich and West (2008) found that cognitive ability was uncorrelated with the biases listed above, they found that increased cognitive ability tended to reduce other cognitive biases: belief bias, matching bias in the Wason selection task, probability matching (instead of maximizing), and denominator neglect. Unfortunately for the friend of the present objection, the GOG error is more like the previous set of cognitive biases than the latter. Stanovich and West argue that smarter people should be less likely to exhibit cognitive bias only "when you tell them what the bias is and what they need to do to avoid it." (2008, p. 690) In other words, intelligence decreases bias only when there is a clearly delineated set of rules one can apply to avoid the bias, and people have knowledge of these rules. Since more intelligent people can apply these rules more easily, they are less prone to the bias. There are no such rules when it comes to biases that are uncorrelated with cognitive ability, and there are no such rules when it comes to the GOG effect. At best, people can avoid falling victim to the GOG effect by actively searching for counterexamples. This option, however, is not available to the epistemological generalists in question, for they believe our justification for endorsing general moral principles comes from reflection on the principles themselves rather than from a search for counterexamples. By generalists' own lights, then, people's tendency to commit the GOG error should be independent of their cognitive ability.

Another possible reason to think philosophers are less likely to fall victim to the GOG effect is that the GOG effect is most likely to occur under time pressure (Meyer, Gelman, and Stilwell, 2010), and philosophers have plenty of time. Philosophers get paid to do philosophy; this allows them a great deal of time for reflection, information-gathering, and other sorts of activities that one might expect to reduce cognitive bias. It would seem to follow that they should be less likely to fall victim to the GOG effect.

We believe the best explanation of why the GOG effect is most likely to occur under time pressure is that people under time pressure have less time to search for counterexamples. But recall, again, that the sort of generalist targeted by our argument is one who *denies that searching for counterexamples is part of properly evaluating a universal generalization*. If so, then, by their own lights, having extra time to search for counterexamples grants them no epistemic advantage over those who are under time pressure.

Of course, having extra time to think might grant some other epistemic advantage to philosophers. As other intuitionists do, Russ Shafer-Landau speaks of "the appearance of intrinsic plausibility that emerges after careful reflection." (2003, p. 260, fn. 9) Likewise, he writes, "That one needs time to see the truth of a proposition is compatible with its being self-evident—some complex propositions require patience and work before an adequate understanding sets in, and it is only such understanding that is claimed to be sufficient for justification." (2003, p. 249) Perhaps this is true. Perhaps lengthy reflection on a principle can allow one a greater understanding of that principle, and perhaps a greater understanding of a principle can give it a glow of intrinsic plausibility that it initially lacked. But it's also likely that any glow of intrinsic plausibility that emerges after a long period of reflection is due to epistemically irrelevant factors, such as mere repetition. In fact, a great deal of psychological literature on what has been called the "truth effect" reveals that people are more inclined to accept a statement the more they hear it repeated, especially when compared to other statements they have not heard repeated.[24] The greater confidence one may gain through time in the belief that one's attraction to a principle is due to its intrinsic plausibility rather than the GOG error may itself be due to the truth effect.

There may still be other reasons to think philosophers are less susceptible to the GOG error that we have overlooked. Nevertheless, the prospects for this hypothesis appear dim. Studies that have been conducted thus far on philosophers' susceptibility to cognitive biases have found that they are not less susceptible. Schwitzgebel and Cushman (2012) found that philosophers' moral judgments about various cases remain susceptible to order effects, while Tobia, Buckwalter, and Stich (2012) found that philosophers' moral judgments can also be influenced by the actor-observer bias (albeit in the opposite direction than is typical). These studies provide direct evidence against the claim that philosophers are immune to cognitive biases like the GOG error.

Objection: The argument against epistemological generalism is self-defeating, as it relies on the principle that people cannot be justified in holding a belief when they are justified in believing that the belief was formed by a process that regularly leads to false beliefs, and they have no other independent grounds that would justify the belief. This is a characteristic universal; therefore, we lack justification for endorsing it.

Reply: This principle is in fact a characteristic universal, and it does enjoy a high degree of intuitive plausibility that could be attributed to the GOG effect. Indeed, the plausibility of the principle is so widely admitted that it is the one internalist condition on justification recognized in some form or another by some of epistemology's most ardent externalists (cf. Nozick, 1981, p. 196; Goldman, 1986, pp. 62–63, 111–112).[25] Nevertheless, there are reasons to accept the principle that are independent of its intrinsic plausibility and the psychological difficulty of rejecting it. In particular, it is supported by a host of particular cases in which we judge that people who learn that their belief is caused by an unreliable process *and not independently justified* but continue to hold that belief are unjustified in doing so, and by the absence of contrary cases. Starting with our judgments about these cases, we can use inference to the best explanation to arrive at the principle.

Although this principle enjoys a great deal of inferential support that is left unscathed by the argument against epistemological generalism, we submit that a version of the argument against generalism would go through even if we were to discard it. If our opponents deny this, we invite them to look directly at the kinds of cases in question. Consider people who believe that a universal moral generalization like "Lying is always wrong" is true solely because they find it intrinsically plausible or difficult to reject. Now imagine they are justified in believing that this belief was formed by an unreliable process. Is it not clear that these people would not be justified in continuing to believe that lying is always wrong? If so, then, far from presupposing the truth of the principle in question, the epistemological argument against generalism seems to provide *support* for it. We conclude, then, that the argument against epistemological generalism is not self-defeating.

Objection: Even characteristic generics admit of genuine counterexamples, so conflicts between moral principles and particular judgments will remain. How are they to be resolved?

Reply: This is not so much an objection, as an invitation to think clearly about the difference between exceptions and counterexamples. In general, exceptions to generics are tolerated if the exceptional member of the kind simply lacks the property in question, and does not have an equally salient, vivid, concrete alternative property instead (see Leslie, 2008, for details). However, in the case of characteristic generics at least, the situation can be somewhat more complicated. Ravens are black, and we are prepared to hold to this characteristic generic even while recognizing the existence of white albino ravens. But we gave up our belief in the characteristic generic to the effect that swans are white after finding a lot of black swans in Australia. We found that having white as one's color was not in fact a characteristic upshot of one's being a swan. We presumably arrived at this by an inductive inference to the best explanation of the widespread existence of black swans in south east and south west Australia. The explanation of this and other facts like inheritance of feather color within the Australian swans was that black swans form a species of swan (Cygnus atratus). From this it followed that having white as one's color is not in fact a characteristic upshot of being a swan.

Likewise, there can be false moral generics, e.g. that sexual acts outside of religiously consecrated marriage are wrong, which can be shown to be false in a similar way. But this is no help to the generalist when it comes to moral epistemology. Quite the contrary; the proper epistemic route to giving up such false characteristic generics begins with contemplation of particulars, e.g. morally unproblematic secular marriages, and goes on to use the best explanation of those particular facts to defeat the original characteristic generic.

In general we defeat a characteristic generic by an induction *from particular cases* which undermines the implied claim of characteristic connection. As far as the prospects for a variant of generalism restricted to moral *generics*, the important thing to note is that the original generic is not to be insulated from the impact of particular *counterexamples* by the mere fact that it is general in form. Rather, the generic stands only so long as it remains the best explanation of the particular data we have at hand.

Section 7: A Way Forward

We have argued that we ultimately have no non-inferential justification for favoring judgments about general moral principles over incompatible judgments about particular cases. Our arguments leave open the possibility, however, that we might have inferential justification for taking the generalist route in response to familiar dilemmas. In the most general terms, we might justifiably favor a general moral principle over an incompatible judgment about a particular case when we have reason to believe our particular judgment is the result of a less reliable process than our judgment about the general principle.

For an example of how this might work, consider the familiar dilemma raised by Peter Singer in his (1972) "Famine, Affluence, and Morality." Singer offers a seemingly innocuous principle: "If it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do it." (1972, p. 231) This principle sounds right. It seems both intrinsically credible and difficult to reject. But this principle implies something counterintuitive about how much money we are obligated to give away to charities that are working to alleviate extreme poverty: "We ought to give until we reach the level of marginal utility—that is, the level at which, by giving more, I would cause as much suffering to myself or my dependents as I would relieve by my gift." (Singer, 1972, p. 241) This in turn seems to imply that it was wrong of you to buy that luxury sedan you bought last year, because you could have bought an economy car instead and donated the money you saved to an effective charity, where it would have been used to protect a large number of people in a developing country from various severe but preventable diseases. People often respond to Singer's argument by pointing out that it was not wrong for them to buy that luxury sedan last year, and so Singer's principle must be wrong.

How could we resolve this familiar dilemma in the generalist's favor? One way would be to argue that Singer's principle is self-evident, but that the proposition that we did nothing wrong in buying a luxury sedan is not. We have argued in this paper that this sort of option is unavailable. If we were to be justified in endorsing the principle because it is self-evident, our attraction to the principle would have to be a result of our adequately understanding the principle. We have good reason to believe, however, that our attraction to Singer's principle—which is equivalent to the claim "Preventing bad things from happening at no net moral cost is always obligatory"—is due instead to the fact that its generic counterpart—"Preventing bad things from happening at no net moral cost is obligatory"—is true.

Another way to resolve the dilemma in the generalist's favor would be to argue that we are more confident that Singer's principle is true than we are that our judgment about the sedan is true, or that the principle has a better claim on seeming true than the claim about the sedan. We have argued in this paper that these moves are also unavailable. These moves are unavailable because we are justified in believing that our belief that Singer's principle is true is the product of a process that regularly produces false beliefs, and we cannot be justified in holding a belief when we are justified in believing that belief to have been the product of an unreliable process and we have no independent justification for it.

A better strategy is to grant that our judgments about particular cases are reliable and then argue that Singer's principle is the result of some kind of inductive process that takes these judgments as input. Either Singer's principle is the best explanation of a wide variety of particular moral facts we take ourselves to know, or else enumerative induction on the basis of these facts leads us to posit it. Even granting that general moral principles can be justified in this way, however, is not enough to justify us in taking the generalist route in the face of familiar dilemmas; since this strategy for justifying Singer's principle assumes the general reliability of our judgments about particular cases, it permits us to reject Singer's principle on the basis of the fact that it conflicts with our judgment about the luxury sedan. In order to avoid this result and support the generalist move, then, an additional condition must be met: We must have independent reason for doubting the reliability of our judgment concerning the permissibility of buying a luxury sedan. We must show that our judgment that buying a luxury sedan was permissible is influenced by some unreliable process—or at least some process that is less reliable than whatever form of inductive inference we used to arrive at Singer's principle. Only if we can attribute our anti-Singerian judgment to something like selfishness or out-group bias—processes which have both proven unreliable in the past—will we have proper justification for favoring Singer's principle over our judgment about buying a luxury sedan.[26]

In conclusion, we suggest that familiar dilemmas be resolved by checking to see whether the principle in question can be supported on the basis of some inductive procedure that takes our judgments about particular cases as input, and then checking to see whether our incompatible judgment about the particular test case is the product of an unreliable process. If these two conditions are met, then the dilemma should be resolved in the generalist's favor. Otherwise, we should proceed as particularists.

Either way, the psychological facts about how we generalize suggest that there is no good reason to privilege intuitions about general moral principles when they conflict with apparent counterexamples.

Notes

1. The authorship on this paper follows the convention according to which A.L. is the primary (first) author, and S.J.L. is the secondary author.
2. The terms "generalism" and "particularism" are usually reserved for metaphysical theses concerning the structure of morality. Nevertheless, the present use is not without precedent; Dancy (2013) refers to the epistemological theses at issue here as "generalist epistemology" and "particularist epistemology." Likewise, in his classic *The Problem of the Criterion*, Roderick Chisholm uses "particularism" to refer to the view that our judgments about what we know in particular cases are epistemically prior to our judgments about what it takes for a belief to count as an instance of knowledge.
3. Most famously, consequentialists like Peter Singer (1972, 1979), Peter Unger (1996), and J.J.C. Smart (1973) have relied on intuitively plausible consequentialist principles to drive significant revision in judgments about particular cases. Capturing the motivation behind the method, Smart writes, "in some moods the general principle of utilitarianism may recommend itself to us so much the more than do more particular moral precepts, precisely because it is so general." (1965, p. 345) Privileging intuitions about general principles is not, however, limited to consequentialists. For instance, Robert Nozick (1974) famously relied on what he took to be the clear existence of moral side constraints on action in order to draw counterintuitive conclusions about the permissibility of economic redistribution. More recently, Judith Jarvis Thomson (2008) seems to argue in part from the intuitive plausibility of general deontological principles to the counterintuitive conclusion that it is wrong for the bystander to turn the train onto the side track in the 'switch' version of Philippa Foot's (1967) trolley problem.
4. Shafer-Landau does not commit himself to this view, but he does recognize it as a plausible move for the self-evidence theorist to make.
5. The terminology is due to Bedke (2010).
6. While Huemer does defend seeming state theory, he does not use it himself to defend substantive general principles of the sort other seeming state theorists might. Huemer limits his epistemological generalism to defending "formal intuitions" such as "If *x* is better than *y* and *y* is better than *z*, then *x* is better than *z*," and "If it is wrong to do *x*, and it is wrong to do *y*, then it is wrong to do both *x* and *y*." (2008, p. 386) For Huemer's own criticisms of generalist epistemology, see his (2005, pp. 166–167) and (2008, pp. 383–387).
7. It is a commonplace of modern moral philosophy that most of its practitioners follow the method of reflective equilibrium in something like the way it was articulated by John Rawls. To the extent that this is true, most philosophers should in principle be open to taking the generalist route in response to familiar dilemmas:

   > People have considered judgments at all levels of generality, from those about particular situations and institutions up through broad standards and first principles to formal and abstract conditions on moral conceptions. One tries to see how people would fit their various convictions into one coherent scheme, each considered conviction whatever its level having a certain initial credibility. By dropping and revising some, by reformulating and expanding others, one supposes that a systematic organization can be found. Although

> in order to get started various judgments are viewed as firm enough to be taken provisionally as fixed points, there are no judgments on any level of generality that are in principle immune to revision. Even the totality of particular judgments are not assigned a decisive role; thus these judgments do not have the status sometimes attributed to judgments of perception in theories of knowledge. (Rawls 1974, p. 8)

Despite the fact that Rawls clearly makes room for the favoring of intuitions about general moral principles over intuitions about particular cases, moral philosophers rarely take this option. On behalf of his colleagues, Shelly Kagan writes, "I think it fair to say that almost all of us trust intuitions about particular cases over general theories, so that given a conflict between a theory—even one that seems otherwise attractive—and an intuitive judgment about a particular case that conflicts with that theory, we will almost always give priority to the intuition." (2001, p. 45) This is not to say that no moral philosophers favor their intuitions about general principles—many do, and to great effect. Nevertheless, surveying the field as a whole, philosophers take the generalist route less often than one might expect.

This is a curious state of affairs. Consequently, one might expect there to be a principled reason why, when intuitions about general principles conflict with intuitions about particular cases, philosophers tend to favor their intuitions about particular cases. According to Kagan, however, no such explanation is forthcoming. While one might claim that our intuitions about particular cases are epistemically privileged in virtue of the fact that they are about a special kind of object—particular cases—Kagan denies this:

> [T]his reassuring answer is itself threatened by the realization that this very distinction between two *kinds* of objects for intuitions may well be misguided. For the fact of the matter, I believe, is that when we react to particular cases we are actually reacting to things of the very same type as when we react to general moral claims. It is easy to lose sight of this, given our common practice—one that I have followed in this paper as well—of saying that we are reacting to *particular* cases. But what we are actually reacting to, I think, are *types* of cases. (p. 61)

This, in turn, casts doubt on the dominance of particularism in moral epistemology:

> If all, or at least most, case specific intuitions are not actually reactions to something concrete and particular at all, then we cannot readily claim that what makes intuition more reliable here is that it is directed at a different kind of object than when we intuitively respond to a general moral claim. In both cases, it seems, what we see is something general.
>
> Of course, there will still be differences in degrees of generality, and it might be that what we should give priority to are our intuitive reactions to the less general rather than to the more general. But this, too, calls out for explanation, and it is not clear what could be said in its defense. (p. 62)

In this paper, we aim not only to undermine epistemological generalism where it exists, but to give epistemological particularists something to say in response to

Kagan. If we are right, the current dominance of epistemological particularism in moral epistemology is justified by the fact that we lack non-inferential justification for endorsing general moral principles when they conflict with intuitions about particular cases. This point remains even if, with Kagan, we regard the judgments about particular cases as judgments about types of cases. The claim would then be that the more general moral claims have the epistemic status they do because of their relations to the more particular moral claims.

8. See, e.g., Lance and Little (2004).
9. For data on snake species, see Russell (1990). The assertion that most individual snakes are non-venomous is based on the assumption that the average population of venomous species is no larger than the average population of non-venomous species. While this assumption may be false, the point here is that the generic "Snakes are venomous" would remain true even if we found out that most snakes are not venomous.
10. The most common formulations of such moral principles are given in gerundive statements, e.g. "Lying is wrong," or related constructions, such as "It is wrong to lie." Some theorists explicitly identify such statements as generic (e.g., see Carlson & Pelletier, 1995, for examples). Even if one takes a more restricted view of what a generic *sentence* is (e.g., Leslie, 2008), this is compatible with recognizing that these gerundive statements have a great deal in common with generic statements, particularly in as much as they tolerate exceptions. (E.g., we might note that they can be reformulated as true bare plural generics without changing anything that is important to moral theorizing: "Acts of lying are wrong.") This paper will be primarily concerned with the sorts of sentiments that lie behind these generalizations, rather than with technical questions concerning syntactic form. (That is, a statement may express a claim with a generic flavor—perhaps by articulating one's underlying generic belief—even if it has different surface syntax.)
11. Leslie et al. also conducted another follow-up experiment, in which participants evaluated the generalizations as before, but were asked afterwards to rephrase the same generalizations in their own words. If participants had mentioned subkinds or restricted domains in their paraphrases, this would have lent support to these deflationary explanations of the GOG effect. As a matter of fact, only 1.6% of paraphrases reflected that participants had interpreted the generalizations to involve some quantifier domain restriction or subkind quantification. Further, these suggestive paraphrases were uncorrelated with endorsement of universals—i.e., participants who provided such paraphrases were no more likely to endorse the universals than those who did not. All in all, these findings make it unlikely that the GOG effect is to be explained by participants quantifying over subkinds or utilizing domain restriction (Leslie et al., 2011, Experiment 2b).
12. Participants rejected the universal and indicated they knew that one sex lacked the property on 44% of trials, and displayed ignorance of the relevant sex-linked information on the remaining 16% of trials.
13. In a final experiment, Leslie et al. showed participants each generalization from the previous experiments in both universal form (e.g., "All Xs are Ys") and existential form (e.g., "Only some Xs are Ys" or "Some Xs are not Ys"). Participants were asked to report on a scale of 1 to 6 whether they agreed more with the universal version of the generalization or the existential version. Where

a score of 6 reflects complete agreement with universals, and a score of 1 reflects complete agreement with existentials, the average score in response to majority characteristic generalizations was 4.61 and the average score in response to the minority characteristic generalizations was 3.56. The first score clearly reflects the presence of a GOG effect for majority characteristic universals. Although the second score was close to the point of indifference between the minority characteristic universals and existentials, it also reflects a greater tendency toward endorsement of minority characteristic universals than non-characteristic universals; mean scores for non-characteristic universals were all below 3. Furthermore, the distribution of scores for minority characteristic generalizations was significantly bimodal—most scores fell closer to the end-points of the scale than the mid-point, reflecting the fact that there were some participants who clearly fell victim to the GOG effect. According to Leslie et al., "participants who preferred the minority characteristic universal did so with confidence." (2011, p. 27)

[14] See, e.g., Staudenmayer, 1975; Wason & Johnson-Laird, 1972. For a review of the literature, see Politzer & Bourmard (2002).

[15] For similar studies, see (e.g.) Byrne (1989), Cummins et al. (1991), De Neys et al. (2002).

[16] Another strand to consider here is that two recent studies suggest that, once people have accepted a generic, they treat that generic as being inferentially powerful, in that they show a default expectation that a given member of the kind will have the property in question—where that expectation does not reduce to their statistical beliefs about how prevalent the property is among members of the kind. For example, Khemlani, Leslie, and Glucksberg (2012) asked people to evaluate whether they thought that an arbitrary member of a kind would have a property—that is, people were asked to judge, e.g., whether Quacky the duck would lay eggs, or (alternatively) whether Quacky the duck would be female. Despite correctly giving comparable prevalence estimates for the percentage of ducks that lay eggs vs. are female, participants were significantly more confident that Quacky lays eggs than they were that Quacky is female. More generally, when participants accepted a background generic (as they did with "Ducks lay eggs"), they were significantly more confident that an individual would have the property in question than they were when they rejected the background generic (as they did with "Ducks are female") *even when their prevalence estimates were controlled for*. Further, Cimpian, Brandone, and Gelman (2010) found that people showed a tendency to accept novel generics (e.g., "Lorches have dangerous feathers") when told that, e.g., only 30% of lorches have dangerous feathers. However, if participants were simply told "Lorches have dangerous feathers" and then asked to estimate the percentage of lorches have such feathers, the majority of participants estimated that 100% of lorches would have the property.

These two studies suggest a certain asymmetry between the conditions under which people will accept generics, and the inferences they are willing to draw from generics. In particular, once they have accepted a generic, they may assume, by default, that any given member of the kind has the property—even though this being the case it not required for the generic to be accepted. This possibility—if correct—would mean that our ability to think of counterexamples to universal claims may itself be further hampered by our acceptance of its generic counterpart. That is, if accepting a generic disposes us to think that a given member of the kind has the property, then this fact itself may limit our ability to come up with counterexamples to the generic's universal counterpart.

17. Of course, it is possible that high working memory capacity improves counterexample retrieval without improving people's ability to use counterexamples in reasoning. Another study helps rule out this possibility. Using verbal report methods, in which participants were instructed to think aloud as they engaged in conditional reasoning, Verschueren, Schaeken, and d'YDewalle (2005) found that people who were higher in working memory capacity were more likely to use counterexamples in conditional reasoning. Complementing previous studies showing that higher working memory capacity leads to improved conditional reasoning, Verschueren et al.'s findings suggest that higher working memory capacity not only leads to improved conditional reasoning and improved counterexample retrieval, but that it leads to improved conditional reasoning via improved counterexample retrieval. This lends further support to the claim that counterexample retrieval is a cognitively demanding process that must compete with and overcome people's initial inclination to accept false universals.
18. The hypothesis that people are susceptible to the GOG effect in part because they fail to retrieve counterexamples to the universal in question also has the advantage of explaining why people are more prone to the GOG error when evaluating *majority* characteristic universals than when evaluating *minority* characteristic universals; presumably, it is easier to retrieve a counterexample to a minority characteristic universal, insofar as there are more of them.
19. Indeed, even those who doubt the existence of moral universals seek principles that can do explanatory work. When Mark Lance and Maggie Little set out to defend the centrality of defeasible generalizations—non-universal generalizations riddled with exceptions—to moral theory, one of their primary concerns is to show that defeasible generalizations can remain explanatory, even when they are subject to counterexample: "Pointing to the fact that an action is a case of lying is explanatory in a way that pointing to surrounding detail is not (and this even though the moral landscape is rife with exceptions to lying's wrong-making status) because lying is defeasibly wrong-making." (2008, p. 71) Likewise, when leading particularist Jonathan Dancy rejects the importance of moral principles to moral theorizing, he is largely concerned to deny that we need general moral principles to *explain* particular moral facts. For instance, Dancy denies that we need to invoke any general principles about the wrongness of killing innocent people in order to explain why there is some reason not to kill a particular person: "The question I want to raise is whether the fact that this feature (that we are causing the death of an unwilling and blameless victim) is functioning as the reason it here is, is in any way to be explained by appeal to the (supposed) fact that it functions in the same way in every case in which it occurs. It seems to me that this feature is the reason it is here quite independently of how it functions elsewhere." (2004, p. 79)
20. That is, unless they are justified in believing that the belief was simultaneously formed by a more fine-grained process that *is* reliable (Sinnott-Armstrong, 2009). We consider the objection from self-evidence theorists that we do employ such a fine-grained process below.

21. Is the principle a generic or a universal? We are inclined to believe it is universal, but of course the thrust of this paper is that such inclinations are not to be trusted! We consider this objection in section 6.
22. For a similar argument that none of our moral beliefs (about general principles *or* particular cases) are non-inferentially justified, see Sinnott-Armstrong (2006, 2009). Although different in a number of respects, the formulation of the present argument takes much of its inspiration from the argument offered in Sinnott-Armstrong (2009).
23. Although Shafer-Landau's principles include a ceteris paribus clause, they are in fact exceptionless principles ascribing *pro tanto* wrongness to every instance of the act-type in question. Shafer-Landau makes this clear in the next chapter of his book when he attacks exception-laden principles as being compatible with particularism, which he does not endorse: "Alternatively, we can understand a property's being generally relevant as its being typically relevant, though in some cases the value ordinarily conveyed by the property's instantiation is either absent or reversed. So, for instance, we might say that beneficence typically is a force for good, though in unusual cases, a person's beneficence either generates no moral credit, or is instead a kind of viciousness. But on this understanding of general relevance, we are right back to particularism." (2003, p. 270)
24. For a review and meta-analysis of the empirical literature on the truth effect, see Dechêne, Stahl, Hansen, and Wänke (2010).
25. We owe these references to Sudduth (2008).
26. To the best of our knowledge, this kind of appeal to second-order beliefs concerning the reliability of the processes underlying our first-order moral judgments was first discussed at length by Walter Sinnott-Armstrong in his (1996) "Moral Skepticism and Justification."

References

Audi, R. (2009). Intuitions, intuitionism, and moral judgment. In J. G. Hernandez (Ed.), *The new intuitionism* (pp. 171–199). New York: Continuum International Publishing Group.

Barzun, J. (1983). *A stroll with William James*. New York: Harper & Row.

Bedke, M. S. (2010). Intuitional epistemology in ethics. *Philosophy Compass*, *5*(12), 1069–1083.

Bloom, P. (2013). *Just babies*. New York: Crown.

Byrne, R. M. J. (1989). Suppressing valid inferences with conditionals. *Cognition*, *31*(1), 61–83.

Carlson, G. N., & Pelletier, F. J. (1995). *The generic book*. Chicago: Chicago University Press.

Chisholm, R. M. (1973). *The problem of the criterion*. Milwaukee: Marquette University Press.

Cimpian, A., Brandone, A. C., & Gelman, S. A. (2010). Generic statements require little evidence for acceptance but have powerful implications. *Cognitive Science*, *34*(8), 1452–1482.

Cummins, D. D, Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, *19*(3), 274–282.

Dancy, J. (2004). *Ethics without principles*. Oxford: Oxford University Press.

Dancy, J. (2013). Moral epistemology. In E. Sosa and M. Steup (Eds.), *A companion to epistemology: second edition*. Oxford: Wiley-Blackwell. Retrieved from Blackwell Reference Online, http://www.blackwellreference.com/subscriber/tocnode.html?id=g9781405139007_chunk_g978140513900743_ss1-11

Daniels, N. (1979). Wide reflective equilibrium and theory acceptance in ethics. *The Journal of Philosophy*, *76*(5), 256–282.

Darley, J., & Schultz, T. R. (1990). Moral rules: Their content and acquisition. *Annual Review of Psychology*, *41*, 525–556.

Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, *14*(2), 238–257.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2002). Causal reasoning and semantic memory retrieval: A test of the semantic memory framework. *Memory & Cognition*, *30*(6), 908–920.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2003). Inference suppression and semantic memory retrieval: every counterexample counts. *Memory & Cognition*, *31*(4), 581–595.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005a). Working memory and counterexample for causal conditionals. *Thinking & Reasoning*, *11*(2), 123–150.

De Neys, W., Schaeken, W., & d'Ydewalle, G. (2005b). Working memory and everyday conditional reasoning: Retrieval and inhibition of stored counterexamples. *Thinking & Reasoning*, *11*(4), 349–381.

Foot, P. (1967). The problem of abortion and the doctrine of the double effect. *Oxford Review*, *5*, 5–15.

Gelman, S. A. (2010). Generics as a window onto young children's concepts. In F. J. Pelletier (Ed.). *Kinds, Things, and Stuff*. New York: Oxford University Press, pp. 100–123.

Goldman, A. I. (1986). *Epistemology and cognition*. Cambridge, MA: Harvard University Press.

Goodman, N. (1954). *Fact, fiction, and forecast*. Cambridge, MA: Harvard University Press.

Hollander, M. A., Gelman, S. A., & Star, J. (2002). Children's interpretation of generic noun phrases. *Developmental Psychology*, *38*, 883–894.

Huemer, M. (2005). *Ethical intuitionism*. New York: Palgrave Macmillan.

Huemer, M. (2007). Compassionate phenomenal conservatism. *Philosophy and Phenomenological Research*, *74*(1), 30–55.

Huemer, M. (2008). Revisionary intuitionism. *Social Philosophy and Policy*, *25*(1), 368–392.

Johnston, M., & Leslie, S. J. (2012). Concepts, analysis, generics and the Canberra plan. *Philosophical Perspectives*, *26*, 113–171.

Jönsson, M. L., & Hampton, J. A. (2006). The inverse conjunction fallacy. *Journal of Memory and Language*, *55*, 317–334.

Kagan, S. (2001). Thinking about cases. *Social Philosophy and Policy*, *18*(2), 44–63.

Khemlani, S., Leslie, S. J., & Glucksberg, S. (2012). Inferences about members of kinds: The generics hypothesis. *Journal of Language and Cognitive Processes*, *27*(6), 887–900.

Khemlani, S. S., & Johnson-Laird, P. N. (2012). Hidden conflicts: Explanations make inconsistencies harder to detect. *Acta Psychologica*, *139*, 486–491.

Kratzer, A. (1991). Conditionals. In A. von Stechow & D. Wunderlich (Eds.), *Semantics: An international handbook of contemporary research* (pp. 651–657). New York: de Gruyter

Lance, M. N., & Little, M. (2004). Defeasibility and the normative grasp of context. *Erkenntnis*, *61*(2/3), 435–455.

Lance, M. N., & Little, M. (2008). From particularism to defeasibility in ethics. In M. Potrč and V. Strahovnik (Eds.), *Challenging moral particularism* (pp. 53–74). New York: Routledge.

Leslie, S. J. (2007). Generics and the structure of the mind. *Philosophical Perspectives*, *21*, 375–403.

Leslie, S. J. (2008). Generics: Cognition and acquisition. *The Philosophical Review*, *117*(1), 1–47.

Leslie, S.J. (2012). Generics articulate default generalizations. *Recherches Linguistiques de Vincennes: New Perspectives on Genericity at the Interfaces*, *41*, 25–45.

Leslie, S. J., & Gelman, S. A. (2012). Quantified statements are recalled as generics. *Cognitive Psychology*, *64*, 186–214.

Leslie, S. J., Khemlani, S., & Glucksberg, S. (2011). All ducks lay eggs: The generic overgeneralization effect. *Journal of Memory and Language*, *65*, 15–31.

Mannheim, B., Gelman, S. A., Escalante, C., Huayhua, M., & Puma, R. (2011). A developmental analysis of generic nouns in Southern Peruvian Quechua. *Language Learning and Development*, *7*(1), 1–23.

McMahan, J. (2013). Moral intuition. In H. LaFollette and I. Persson (Eds.), *Blackwell guide to ethical theory, second edition* (pp. 103–120). Oxford: Wiley-Blackwell.

Meyer, M., Gelman, S. A., & Stilwell, S. M. (2010). Generics are a cognitive default: Evidence from sentence processing. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 913–918). Boston, MA: Cognitive Science Society.

Nisbett, R. E., Krantz, D. H., Jepson, C., and Kunda, Z. (1983), The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*, 339–363.

Nozick, R. (1974). *Anarchy, state, and utopia*. New York: Basic Books, Inc.

Nozick, R. (1981). *Philosophical explanations*. Cambridge, MA: Harvard University Press.

Politzer, G., & Bourmaud, G. (2002). Deductive reasoning from uncertain conditionals. *British Journal of Psychology*, *93*, 345–381.

Prasada, S., & Dillingham, E. M., (2006). Principled and statistical connections in common sense conception. *Cognition*, *99*, 73–112.

Prasada, S., & Dillingham, E. M. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science*, *33*, 401–48.

Prasada, S., Khemlani, S., Leslie, S. J., & Glucksberg, S. (2013). Conceptual distinctions amongst generics. *Cognition*, *126*, 405–22.

Rawls, J. (1951). Outline of a decision procedure for ethics. *The Philosophical Review*, *60*(2), 177–197.

Rawls, J. (1974). The independence of moral theory. *Proceedings and Addresses of the American Philosophical Association*, *48*, 5–22

Russell, F. E. (1990). When a snake strikes. *Emergency Medicine*, *22*(12), 33–34, 47–40, 43.

Schwitzgebel, E., & Cushman, F. (2012). Expertise in moral reasoning? Order effects on moral judgment in professional philosophers and non-philosophers. *Mind & Language*, *27*(2), 135–153.

Shafer-Landau, R. (2003). *Moral realism: A defense*. Oxford: Oxford University Press.

Singer, P. (1972). Famine, affluence, and morality. *Philosophy & Public Affairs*, *1*(3), 229–243.

Singer, P. (1974). Sidgwick and reflective equilibrium. *The Monist*, *58*(3), 490–517.

Singer, P. (1979). *Practical ethics*. Cambridge, UK: Cambridge University Press.

Sinnott-Armstrong, W. (1996). Moral skepticism and justification. In W. Sinnott-Armstrong & M. Timmons (Eds.), *Moral knowledge? New readings in moral epistemology*. Oxford: Oxford University Press.

Sinnott-Armstrong, W. (2006). Moral intuitionism meets empirical psychology. In T. Horgan & M. Timmons (Eds.), *Meta-ethics after Moore* (pp. 339–336). New York: Oxford University Press.

Sinnott-Armstrong, W. (2009). An empirical challenge to moral intuitionism. In J. G. Hernandez (Ed.), *The new intuitionism* (pp. 11–28). New York: Continuum International Publishing Group.

Sloman, S. A. (1993). Feature-based induction. *Cognitive Psychology*, *25*, 231–280.

Sloman, S. A. (1998). Categorical inference is not a tree: The myth of inheritance hierarchies. *Cognitive Psychology*, *35*, 1–33.

Smart, J. J. C. (1965). The methods of ethics and the methods of science. *The Journal of Philosophy*, *62*(13), 344–349.

Smart, J. J. C. (1973). An outline of a system of utilitarian ethics. In J. J. C. Smart & B. Williams (Eds.), *Utilitarianism: For and against* (pp. 3–74). Cambridge, UK: Cambridge University Press.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology*, *94*(4), 672–695.

Staudenmayer, H. (1975). Understanding conditional reasoning with meaningful propositions. In R. Falmagne (Ed.), *Reasoning: Representation and process in children and adults*. Hillsdale, N.J.: Erlbaum.

Sudduth, M. (2008). Defeaters in epistemology. In J. Fieser and B. Dowden (Eds.), *Internet encyclopedia of philosophy*. Retrieved from http://www.iep.utm.edu/ep-defea/

Tardif, T., Gelman, S. A., Fu, X., & Zhu, L. (2011). Acquisition of generic noun phrases in Chinese: Learning about lions without an "-S". *Journal of Child Language*, *30*, 1–32.

Thomson, J. J. (2008). Turning the trolley. *Philosophy & Public Affairs*, *36*(4), 359–374.

Tobia, K. P., Buckwalter, W., & Stich, S. (2012). Moral intuitions: Are philosophers experts? *Philosophical Psychology*, *iFirst*, 1–10.

Unger, P. (1996). *Living high and letting die*. New York: Oxford University Press.

Van Bavel, J. J., Packer, D. J., Haas, I. J., & Cunningham, W. A. (2012). The importance of moral construal: Moral versus non-moral construal elicits faster, more extreme, universal evaluations of the same actions. *PLOS ONE*, *7*(11), e48693.

Verschueren, N., Schaeken, W., & d'Ydewalle, G. (2005) Everyday conditional reasoning: A working memory-dependent tradeoff between counterexample and likelihood use. *Memory &n Cognition*, *33*(1), 107–119.

Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content*. Cambridge, MA: Harvard University Press.

West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology*, *103*(3), 506–519.