THIS PDF FILE
FOR PROMOTIONAL USE ONLY

2 Framing Moral Intuitions

Walter Sinnott-Armstrong

If you think that affirmative action is immoral, and I disagree, then it is hard to imagine how either of us could try to convince the other without appealing at some point either implicitly or explicitly to some kind of moral intuition. The same need for intuition arises in disputes about other moral issues, including sodomy, abortion, preventive war, capital punishment, and so on. We could never get started on everyday moral reasoning about any moral problem without relying on moral intuitions. Even philosophers and others who officially disdain moral intuitions often appeal to moral intuitions when refuting opponents or supporting their own views. The most sophisticated and complex arguments regularly come down to: "But surely *that* is immoral. Hence, . . ." Without some move like this, there would be no way to construct and justify any substantive moral theory.[1] The importance of moral theory and of everyday moral reasoning thus provides lots of reasons to consider our moral intuitions carefully.

Moral Intuitions

I define a "moral intuition" as a strong immediate moral belief.[2] "Moral" beliefs are beliefs that something is morally right or wrong, good or bad, virtuous or vicious, and so on for other moral predicates. Moral beliefs are "strong" when believers feel confident and do not give them up easily. Moral beliefs are "immediate" when the believer forms and holds them independent of any process of inferring them from any other belief either at the time when the belief originated or during the later times when the belief is maintained. Moral intuitions in this sense might arise after reflection on the facts of the situation. They might result from moral appearances that are not full beliefs. Nonetheless, they are not inferred from those facts or appearances. The facts only specify which case the intuition is about. The appearances merely make acts seem morally right or wrong,

and so on. People do not always believe that things really are as they appear, so moral belief requires an extra step of endorsing the appearance of this case. When this extra step is taken independent of inference, and the resulting belief is strong, the resulting mental state is a moral intuition.

In this minimal sense, most of us have some moral intuitions. We can react immediately even to new cases. Sometimes I ask students, for example, whether it is morally wrong to duck to avoid an arrow when the arrow will then hit another person (Boorse & Sorensen, 1988). Most students and others who consider such cases for the first time quickly form strong opinions about the moral wrongness of such acts, even though they cannot cite any principle or analogy from which to infer their moral beliefs.

In addition to *having* moral intuitions, most of us think that our own moral intuitions are *justified*. To call a belief "justified" is to say that the believer ought to hold that belief as opposed to suspending belief, because the believer has adequate epistemic grounds for believing that it is true (at least in some minimal sense). Our moral intuitions do not seem arbitrary to us. It seems to us as if we ought to believe them. Hence, they strike us as justified.

Moral Intuitionism

The fact that our moral intuitions *seem* justified does not show that they really *are* justified. Many beliefs that appear at first sight to be justified turn out after careful inspection to be unjustified. To determine whether moral beliefs really are justified, we need to move beyond psychological description to the normative epistemic issue of how we ought to form moral beliefs.

There are only two ways for moral intuitions or any other beliefs to be justified:

A belief is justified *inferentially* if and only if it is justified only because the believer is able to infer it from some other belief.
A belief is justified *noninferentially* if and only if it is justified independent of whether the believer is able to infer it from any other belief.

Whether a belief is justified inferentially or noninferentially depends not on whether the believer *actually* bases the belief in an actual inference but instead on whether the believer is *able* to infer that belief from other beliefs.

A moral intuition might be justified inferentially. What makes it a moral intuition is that it is not actually based on an *actual* inference. What makes it justified inferentially is that its epistemic status as justified depends on the believer's *ability* to infer it from some other belief. People often form beliefs immediately without actual inference, even though they are able to justify those beliefs with inferences from other beliefs if the need arises. If they are justified only because of this ability to infer, then these moral intuitions are justified inferentially.

However, if every moral belief were justified inferentially, a regress would arise: if a believer needs to be able to infer a moral belief from some other belief, the needed inference must have premises. Either none or some of those premises are moral. If none of the premises is moral, then the inference could not be adequate to justify its moral conclusion.[3] On the other hand, if even one of the premises is moral, then it would have to be justified itself in order for the inference to justify its conclusion. If this moral premise is also justified inferentially, then we would run into the same problem all over again. This regress might go on infinitely or circle back on itself, but neither alternative seems attractive. That's the problem.

To stop this regress, some moral premise would have to be justified noninferentially. *Moral skeptics* argue that no moral belief is justified noninferentially, so no moral belief is justified. To avoid skepticism, *moral intuitionists* claim that some moral intuitions are justified noninferentially. Moral intuitionists do not only claim that some moral beliefs are justified apart from any actual inference. That would not be enough to stop the skeptical regress. To avoid skepticism, moral intuitionists need to claim that some moral beliefs are justified independent of the believer's ability to infer those moral beliefs from any other beliefs.

A variety of moral intuitionists do make or imply this claim. First, some *reliabilists* claim that a moral belief (or any other belief) is justified whenever it results from a process that is in fact reliable, even if the believer has no reason at all to believe that the process is reliable (Shafer-Landau, 2003). If so, and if some reliable processes are independent of inferential ability, then some moral beliefs are justified noninferentially. Another kind of moral intuitionism claims that some moral beliefs are justified only because they appear or seem true and there is no reason to believe they are false (Tolhurst, 1990, 1998). If moral appearances or seemings are not endorsed, then they are not beliefs, so they cannot serve as premises or make the believer able to infer the moral belief. Such *experientialists*, thus, also claim that some moral beliefs are justified noninferentially. Third,

reflectionists admit that moral intuitions are justified only if they follow reflection that involves beliefs about the subject of the intuition, but they deny that the believer needs to infer or even be able to infer the moral beliefs from those other beliefs in order for the moral belief to be justified (Audi, 2004). If so, the moral believer is justified noninferentially. Since moral intuitionism as I define it is endorsed by these and other prominent moral philosophers, I cannot be accused of attacking a straw man.

This kind of moral intuitionism is openly normative and epistemic. It specifies when moral beliefs are justified—when believers ought to hold them. It does not merely describe how moral beliefs are actually formed. Hence, this normative epistemic kind of moral intuitionism is very different from the descriptive psychological theory that Jonathan Haidt calls "social intuitionism" (Haidt, 2006). One could adopt Haidt's social intuitionism and still deny moral intuitionism as I define it. Or one could deny Haidt's social intuitionism and yet accept moral intuitionism under my definition. They are independent positions.

The kind of moral intuitionism that will concern me here is the normative epistemic kind because that is what is needed to stop the skeptical regress. Even if Haidt is right about how moral beliefs are *formed*, that by itself will not address the normative issue of whether or how moral beliefs can be *justified*. To address that issue, we need to ask whether the normative epistemic kind of moral intuitionism is defensible.

The Need for Confirmation

It is doubtful that psychological research by itself could establish any positive claim that a belief *is* justified. Nonetheless, such a claim presupposes certain circumstances whose denial can undermine it. By denying such circumstances, psychological research might thus establish negative conclusions about when or how moral beliefs are *not* justified (where this merely denies that they ought to be believed and does not make the positive claim that they ought not to be believed). For example, suppose I believe that I am next to a pink elephant, and I know that I believe this only because I took a hallucinogenic drug. This fact about the actual origin of my belief is enough to show that my belief is not justified. My belief in the elephant might be true, and I might have independent ways to confirm that it is true. I might ask other people, take an antidote to the hallucinogen, or feel the beast (if I know that the drug causes only visual but not tactile illusions). Still, I am not justified without some such confirmation. Generally, when I know that my belief results from a process that is likely to lead to

error, then I need some confirmation in order to be justified in holding that belief.

Hallucinogenic drugs are an extreme case, but the point applies to everyday experiences as well. If I am standing nearby and have no reason to believe that the circumstances are abnormal in any way, then I seem justified in believing that someone is under six feet tall simply by looking without inferring my belief from any other belief. In contrast, if a stranger is too far away and/or surrounded by objects of unknown or unusual size, and if my vision is all that makes me believe that he is under six feet tall, then my belief will often be false, so this process is unreliable. Imagine that I see him five hundred yards away next to a Giant Sequoia tree, and he looks as if he is under six feet tall. This visual experience would not be enough by itself to make me justified in believing that he is under six feet tall. Of course, I can still be justified in believing that this stranger is under six feet tall if I confirm my belief in some way, such as by walking closer or asking a trustworthy source. However, if I do not and cannot confirm my belief in any way, then I am not justified in holding this belief instead of suspending belief while I wait for confirmation.

The kinds of confirmation that work make me able to justify my belief by means of some kind of inference. If I ask a trustworthy source, then I can use a form of inference called "appeal to authority". If I walk closer to the stranger, then I can infer from my second-order belief that I am good at assessing heights from nearby. Similarly, if I touch the pink elephant, then I can infer from my background belief that my senses are usually accurate when touch agrees with sight. And so on for other kinds of confirmation. Since confirmation makes me able to infer, when I need confirmation, I need something that gives me an ability to infer. In short, I need inferential confirmation.

We arrive, therefore, at a general principle:

If the process that produced a belief is not reliable in the circumstances, and if the believer ought to know this, then the believer is not justified in forming or holding the belief without inferential confirmation.

This principle probably needs to be qualified somehow, but the basic idea should be clear enough: a need for confirmation and, hence, inference is created by evidence of unreliability.

This general principle is not about moral beliefs in particular, but it does apply to moral beliefs among others. When it is restricted to moral beliefs, its instance can serve as the first premise in *the master argument*:

(1) If our moral intuitions are formed in circumstances where they are unreliable, and if we ought to know this, then our moral intuitions are not justified without inferential confirmation.
(2) If moral intuitions are subject to framing effects, then they are not reliable in those circumstances.
(3) Moral intuitions are subject to framing effects in many circumstances.
(4) We ought to know this—that is, (3).
(5) Therefore, our moral intuitions in those circumstances are not justified without inferential confirmation.

I just argued for the general principle that implies Premise 1. What remains is to argue for the rest of the premises.

What Are Framing Effects?

Premise 2 says that framing effects bring unreliability. This premise follows from the very idea of framing effects. Many different kinds of phenomena have been labeled framing effects (for a typology, see Levin, Schneider, & Gaeth, 1998). What I have in mind are effects of wording and context on moral belief.

A person's belief is subject to a *word* framing effect when whether the person holds the belief depends on which words are used to describe what the belief is about. Imagine that Joseph would believe that Marion is fast if he is told that she ran one hundred meters in ten seconds, but he would not believe that she is fast (and would believe that she is not fast and is slow) if he is told that it took her ten seconds to run one hundred meters (or that it took her ten thousand milliseconds to run one hundred meters). His belief depends on the words: "ran" versus "took her to run" (or "seconds" vs. "milliseconds"). This belief is subject to a word framing effect.

Whether Marion is fast can't depend on which description is used. Moreover, she cannot be both fast and slow (relative to the same contrast class). At least one of Joseph's beliefs must be false. He gets it wrong either when his belief is affected by one of the descriptions or when it is affected by the other. In this situation on this topic, then, he cannot be reliable in the sense of having a high probability of true beliefs. If your car started only half of the time, it would not be reliable. Similarly, Joseph is not reliable if at most half of his beliefs are true. That is one way in which framing effects introduce unreliability.

The other kind of framing effect involves *context*. Recall the man standing next to a giant sequoia tree. In this context, the man looks short.

However, if the man were standing next to a bonsai tree, he might look tall. If Josephine believes that the man is short when she sees the man in the first context, but she would believe that the man is tall if she saw the man in the second context, then Josephine's belief is subject to a context framing effect.

A special kind of context framing effect involves *order*. Imagine that Josephine sees the man both next to a sequoia and also next to a bonsai, but her belief varies depending on the order in which she sees these scenes. If she sees the man next to the sequoia first, then she continues to believe that the man is short even after she sees the man next to the Bonsai. If she sees the man next to the bonsai first, then she continues to believe that the man is tall even after she sees the man next to the sequoia. First impressions rule. The order affects the context of her belief, so, again, Josephine's belief is subject to a context framing effect.

In both cases, at least one of Josephine's beliefs must be false. The man cannot be both short and tall (for a man). Hence, Josephine's beliefs on this topic cannot be reliable, since she uses a process that is inaccurate at least half the time. Thus, context framing effects also introduce unreliability.

The point applies as well to moral beliefs. Suppose your friend promises to drive you to the airport at an agreed time. When the time arrives, he decides to go fishing instead, and you miss your flight. His act could be described as breaking his promise or as intentionally failing to keep his promise, but how his act is described cannot affect whether his act is morally wrong. It morally wrong for him to break his promise in these circumstances if and only if it is also morally wrong for him to intentionally fail to keep his promise in these circumstances. What is morally wrong is not affected by such wording.

It is also not affected by the context of belief. Imagine that ten years later you tell me about your friend's failure. Then I form a moral belief about your friend's failure. Whether my belief is correct depends on what happened at the earlier time, not at the later time when I form my belief. My later context cannot affect any of the factors (such as the act's circumstances or consequences and the agent's beliefs or intentions) that determine whether your friend's act was morally wrong. Of course, the context of the action does affect its moral wrongness. If your friend fails to drive you to the airport because he needs to take his child to the hospital to save her life, then his failure to keep his promise is not morally wrong. However, that is the *agent's* context. The *believer's* context, in contrast, does not affect moral wrongness. If it is morally wrong for your friend to go fishing in the context in which he went fishing, then anyone who

forms a moral belief about that act should judge that the act is morally wrong regardless of the context from which the believer views the act. If the moral wrongness of an act did vary with the believer's context, we could never say whether any act is morally wrong, because there are so many different believers in so many different contexts.

Since wording and context of belief do *not* affect what is morally wrong, if wording or context of belief *does* affect moral beliefs about what is morally wrong, then those moral beliefs will often be incorrect. Moral beliefs that vary in response to factors that do not affect truth—such as wording and belief context—cannot reliably track the truth. Unreliability comes in degrees, but the point still holds: moral beliefs are unreliable to the extent that they are subject to framing effects.

Framing Effects on Moral Intuitions

The crucial question now asks: To what extent *are* moral intuitions subject to framing effects? The third premise in the master argument claims that moral intuitions are subject to framing effects in many circumstances. To determine whether this premise is true, we need to determine the extent to which moral judgments vary with framing. Here is where we need empirical research.

Kahneman and Tversky

Framing effects were first explored by Tversky and Kahneman (1981). In a famous experiment, they asked some subjects this question:

> Imagine that the U.S. is preparing for an outbreak of an unusual Asian disease which is expected to kill 600 people. Two alternative programs to fight the disease, A and B, have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows: If program A is adopted, 200 people will be saved. If program B is adopted, there is a 1/3 probability that 600 people will be saved, and a 2/3 probability that no people will be saved. Which of the two programs would you favor? (p. 453)

The same story was told to a second group of subjects, but these subjects had to choose between these programs:

> If program C is adopted, 400 people will die. If program D is adopted, there is a 1/3 probability that nobody will die and a 2/3 probability that 600 people will die. (p. 453)

It should be obvious that programs A and C are equivalent, as are programs B and D. However, 72% of the subjects who chose between A and B favored

A, but only 22% of the subjects who chose between C and D favored C. More generally, subjects were risk averse when results were described in positive terms (such as "lives saved") but risk seeking when results were described in negative terms (such as "lives lost" or "deaths").

The question in this experiment was about choices rather than moral wrongness. Still, the subjects were not told how the policies affect them personally, so their choices seem to result from beliefs about which program is morally right or wrong. If so, the subjects had different moral beliefs about programs A and C than about programs B and D. The only difference between the pairs is how the programs are described or framed. Thus, descriptions seem to affect these moral beliefs. Descriptions cannot affect what is really morally right or wrong in this situation. Hence, these results suggest that such moral beliefs are unreliable.

Moral intuitionists could respond that moral intuitions are still reliable when subjects have consistent beliefs after considering all relevant descriptions. It is not clear that adding descriptions or adding more thought removes framing effects. (I will discuss this below.) In any case, moral believers would still need to know that their beliefs are consistent and that they are aware of all relevant descriptions before they could be justified in holding moral beliefs. That would make them able to confirm their moral beliefs, so this response would not undermine the main argument, which concludes only that moral believers need confirmation for any particular moral belief.

To see how deeply this point cuts, consider Quinn's argument for the traditional doctrine of doing and allowing, which claims that stronger moral justification is needed for doing or causing harm than for merely allowing harm to happen. When the relevant harm is death, this doctrine says, in effect, that killing is worse than letting die. In support of this general doctrine, Quinn appeals to moral intuitions of specific cases:

> In Rescue I, we can save either five people in danger of drowning at one place or a single person in danger of drowning somewhere else. We cannot save all six. In Rescue II, we can save the five only by driving over and thereby killing someone who (for an unspecified reason) is trapped on the road. If we do not undertake the rescue, the trapped person can later be freed. (Quinn 1993, p. 152; these cases derive from Foot, 1984)

Most people judge that saving the five is morally wrong in Rescue II but not in Rescue I. Why do they react this way? Quinn assumes that these different intuitions result from the difference between killing and letting die or, more generally, between doing and allowing harm. However, Horowitz uses a different distinction (between gains and losses) and a

different theory (prospect theory from Kahneman & Tversky, 1979) to develop an alternative explanation of Quinn's moral intuitions:

> In deciding whether to kill the person or leave the person alone, one thinks of the person's being alive as the *status quo* and chooses this as the neutral outcome. Killing the person is regarded as a negative deviation. . . . But in deciding to save a person who would otherwise die, the person being dead is the *status quo* and is selected as the neutral outcome. So saving the person is a positive deviation. . . . (Horowitz, 1998, pp. 377–378)

The point is that we tend to reject options that cause definite negative deviations from the status quo. That explains why most subjects rejected program C but did not reject program A in the Asian disease case, despite the equivalence between those programs. It also explains why we think that it is morally wrong to "kill" in Rescue II but is not morally wrong to "not save" in Rescue I, since killing causes a definite negative deviation from the status quo. This explanation clearly hinges on what is taken to be the status quo, which in turn depends on how the options are described. Quinn's story about Rescue I describes the people as already "in danger of drowning," whereas the trapped person in Rescue II can "later be freed" if not for our "killing" him. These descriptions affect our choice of the neutral starting point. As in the Asian disease cases, our choice of the neutral starting point then affects our moral intuitions.

Horowitz's argument leaves many ways for opponents to respond. Some moral intuitionists argue that, even if the difference between gains (or positive deviations) and losses (or negative deviations) does explain our reactions to Quinn's cases, this explanation does not show that our moral intuitions are incoherent or false or even arbitrary, as in the Asian disease case. Horowitz claims, "I do not see why anyone would think the distinction [between gains and losses] is morally significant, but perhaps there is some argument I have not thought of" (Horowitz, 1998, p. 381). As Mark van Roojen says, "Nothing in the example shows anything wrong with treating losses from a neutral baseline differently from gains. Such reasoning might well be appropriate where framing proceeds in a reasonable manner" (Van Roojen, 1999, p. 854).[4] Indeed, Frisch (1993) found that subjects who were affected by frames often could give justifications for differentiating the situations so described. Nonetheless, the framing also "might well" *not* be reasonable, so there still might be a *need* for some reason to believe that the framing is reasonable. This need produces the epistemological dilemma: if there is *no* reason to choose one baseline over the other, then our moral intuitions seem arbitrary and unjustified. If there *is* a reason to choose one baseline over the other, then either we have access

to that reason or we do not. If we have access to the reason, then we are able to draw an inference from that reason to justify our moral belief. If we do not have access to that reason, then we do not seem justified in our moral belief. Because framing effects so often lead to incoherence and error, we cannot be justified in trusting a moral intuition that relies on framing effects unless we at least can be aware that this intuition is one where the baseline is reasonable. Thus, Horowitz's explanation creates serious trouble for moral intuitionism whenever framing effects could explain our moral intuitions.

A stronger response would be to show that prospect theory is not the best explanation of our reactions to Quinn's cases.[5] Kamm (1998a) argues that the traditional distinction between doing and allowing harm, rather than prospect theory's distinction between gains and losses, is what really drives our intuitions in these cases. These distinctions overlap in most cases, but we can pull them apart in test cases where causing a harm prevents a greater loss, such as this one:

> Suppose we frame rescue 2 so that five people are in excellent shape but need a shot of a drug, the last supply of which is available only now at the hospital, to prevent their dying of a disease that is coming into town in a few hours. Then not saving them would involve losses rather than no-gains. We still should not prevent these five losses of life by causing one loss in this case. So even when there is no contrast between a loss and no-gain in a case, we are not permitted to do what harms (causes a foreseen loss) in order to aid (by preventing a loss). (Kamm, 1998a, p. 477)

Here a failure to save the five is supposed to involve losses to the five, because they are alive and well at present, so the baseline is healthy life. There are, however, other ways to draw the baseline. The disease is headed for town, so the five people are doomed to die if they do not get the drug (just as a person is doomed when an arrow is headed for his heart, even if the arrow has not struck yet). That feature of the situation might lead many people to draw the baseline at the five people being dead. Then not saving them would involve no-gains rather than losses, contrary to Kamm's claim. Thus, prospect theory can explain why people who draw such a baseline believe that we should not cause harm to save the five in this case. Kamm might respond that the baseline was not drawn in terms of who is doomed in the Asian flu case. (Compare her response to Baron at Kamm 1998a, p. 475.) However, prospect theory need not claim that the baseline is always drawn in the same way. People's varying intuitions can be explained by variations in where they draw the baseline, even if they have no consistent reason for drawing it where they do. Thus, Horowitz's explanation does seem to work just fine in such cases.[6]

Psychologists might raise a different kind of problem for Horowitz's argument. Framing effects in choices between risks do not always carry over into choices between definite effects, and they get weaker in examples with smaller groups, such as six hundred people versus six people (Petrinovich & O'Neill, 1996, pp. 162–164). These results together suggest that special features of Asian disease cases create the framing effects found by Kahneman and Tversky. Those features are lacking from Quinn's cases, which do not involve probabilities or large numbers. This asymmetry casts doubt on Horowitz's attempt to explain our reactions to Quinn's cases in the same way as our reactions to Asian disease cases.

Finally, some opponents might respond that Horowitz's claim applies only to the doctrine of doing and allowing, and not to other moral intuitions. However, the doctrine of doing and allowing is neither minor nor isolated. It affects many prominent issues and is strongly believed by many philosophers and common people, who do not seem to be able to infer it from any other beliefs. Similar framing effects are explained by prospect theory in other cases involving fairness in prices and tax rates (Kahneman, Knetsch, & Thaler, 1986) and future generations (Sunstein, 2004, 2005) and other public policies (Baron, 1998). There are still many other areas of morality, but, if moral intuitions are unjustified in these cases, doubts should arise about a wide range of other moral intuitions as well.

To see how far framing effects extend into other moral intuitions, we need to explore whether framing effects arise in different kinds of moral conflicts, especially moral conflicts without probabilities or large numbers. Then we need to determine the best explanation of the overall pattern of reactions. This project will require much research. There are many studies of framing effects outside morality, especially regarding medical and economic decisions. (Kühberger, 1998, gives a meta-analysis of 248 papers.) However, what we need in order to assess the third premise of the master argument are studies of framing effects in moral judgments in particular. Luckily, a few recent studies do find framing in a wider array of moral intuitions.

Petrinovich and O'Neill

Petrinovich and O'Neill (1996) found framing effects in various trolley problems. Here is their description of the classic side-track trolley case:

> A trolley is hurtling down the tracks. There are five innocent people on the track ahead of the trolley, and they will be killed if the trolley continues going straight ahead. There is a spur of track leading off to the side. There is one innocent person on that spur of track. The brakes of the trolley have failed and there is a switch that

can be activated to cause the trolley to go to the side track. You are an innocent bystander (that is, not an employee of the railroad, etc.). You can throw the switch, saving five innocent people, which will result in the death of the one innocent person on the side track. What would you do? (p. 149)

This case differs from Rescues I–II in important respects. An agent who saves the five and lets the one drown in Rescue I does not cause the death of the one. That one person would die in Rescue I even if nobody were around to rescue anyone. In contrast, if nobody were around to throw the switch in the side-track trolley case, then the one person on the side track would not be harmed at all. Thus, the death of the one is caused by the act of the bystander in the side-track trolley case but not in Rescue I. In this respect, the side-track trolley case is closer to Rescue II. It is then surprising that, whereas most people agree that it *is* morally wrong to kill one to save five in Rescue II, most subjects say that it is *not* morally wrong to throw the switch in the side-track trolley case.

The question raised by Petrinovich and O'Neill is whether this moral intuition is affected by wording. They asked 387 students in one class and 60 students in another class how strongly they agreed or disagreed with given alternatives in twenty-one variations on the trolley case. Each alternative was rated on a 6-point scale: "strongly agree" (+5), "moderately agree" (+3), "slightly agree" (+1), "slightly disagree" (–1), "moderately disagree" (–3), "strongly disagree" (–5).[7]

The trick lay in the wording. Half of the questionnaires used "kill" wordings so that subjects faced a choice between (1) ". . . throw the switch which will result in the death of the one innocent person on the side track . . ." and (2) ". . . do nothing which will result in the death of the five innocent people . . ."). The other half of the questionnaires used "save" wordings, so that subjects faced a choice between (1*) ". . . throw the switch which will result in the five innocent people on the main track being saved . . ." and (2*) ". . . do nothing which will result in the one innocent person being saved . . ."). These wordings do not change the facts of the case, which were described identically before the question is posed.

The results are summarized in table 2.1 (from Petrinovich & O'Neill, 1996, p. 152).

The top row shows that the average response was to agree slightly with action (such as pulling the switch) when the question was asked in the save wording but then to disagree slightly with action when the question was asked in the kill wording.

These effects were not due to only a few cases: "Participants were likely to agree more strongly with almost any statement worded to Save than

Table 2.1
Means and standard deviations (in parentheses) of participants' levels of agreement with action and inaction as a function of whether the questions incorporating action and inaction were framed in a kill or save wording[a]

	Saving Wording	Killing Wording
Action	0.65	–0.78
	(0.93)	(1.04)
Inaction	0.10	–1.35
	(1.04)	(1.15)

[a]Positive mean values in the table indicate agreement, and negative values indicate disagreement.
Source: Petrinovich & O'Neill, 1994, p. 152.

one worded to Kill." Out of 40 relevant questions, 39 differences were significant. The effects were also not shallow: "The wording effect . . . accounted for as much as one-quarter of the total variance, and on average accounted for almost one-tenth when each individual question was considered." Moreover, wording affected not only strength of agreement (whether a subject agreed slightly or moderately) but also whether subjects agreed or disagreed: "the Save wording resulted in a greater likelihood that people would absolutely agree" (Petrinovich & O'Neill, 1996, p. 152).

What matters to us, of course, is that these subjects gave different answers to the different questions even though those questions were asked about the same case. The facts of the case—consequences, intentions, and so on—did not change. Nor did the options: throwing the switch and doing nothing. All that varied was the wording of the dependent clause in the question. That was enough to change some subjects' answers. However, that wording cannot change what morally ought to be done. Thus, their answers cannot track the moral truth.

Similar results were found in a second experiment, but this time the order rather than the wording of scenarios was varied. One hundred eighty-eight students were asked how strongly they agreed or disagreed (on the same scale of +5 to –5) with each of the alternatives in the moral problems on one form. There were three pairs of forms.

Form 1 posed three moral problems. The first is the side-track trolley problem. In the second, the only way to save five dying persons is to scan the brain of a healthy individual, which would kill that innocent person. In the third, the only way to save five people is to transplant organs from a healthy person, which would kill that innocent person. All of the options

were described in terms of who would be saved. Form 1R posed the same three problems in the reverse order: transplant, then scan, then side-track. Thirty students received Form 1, and 29 students received Form 1R.

The answers to Form 1 were not significantly different from the answers to Form 1R, so there was no evidence of any framing effect. Of course, that does not mean that there was no framing effect, just that none was found in this part of the experiment.

A framing effect was found in the second part of the experiment using two new forms: 2 and 2R. Form 2 began with the trolley problem where the only way to save the five is to pull a switch. In the second moral problem on Form 2, "You can push a button which would cause a ramp to go underneath the train; the train would jump onto tracks on the bridge and continue, saving the five, but running over the one" (Petrinovich & O'Neill, 1996, p. 156). In the third problem on Form 2, the only way to stop the trolley from killing the five is to push a very large person in front of the trolley. All of the options were described in terms of who would be saved. Form 2R posed the same three problems in the reverse order: Person, then Button, then Trolley. 30 students received Form 2, and 29 received Form 2R.

The results of this part of the experiment are summarized in their table 3 and figure 2 (Petrinovich & O'Neill 1996, pp. 157–158; see table 2.2 and figure 2.1.)

Participants' agreement with action in the Trolley and Person dilemmas were significantly affected by the order. Specifically, "People more strongly approved of action when it appeared first in the sequence than when it appeared last" (Petrinovich & O'Neill, 1996, p. 157). The order also significantly affected participants' agreement with action in the Button dilemma (whose position in the middle did not change when the order changed). Specifically, participants approved more strongly of action in the Button dilemma when it followed the Trolley dilemma than when it followed the Person dilemma.

Why were such framing effects found with Forms 2 and 2R but not with Forms 1 and 1R? Petrinovich and O'Neill speculate that the dilemmas in Forms 1 and 1R are so different from each other that participants' judgments on one dilemma does not affect their judgments on the others. When dilemmas are more homogeneous, as in Forms 2 and 2R, participants who already judged action wrong in one dilemma will find it harder to distinguish that action from action in the other dilemmas, so they will be more likely to go along with their initial judgment, possibly just in order to maintain coherence in their judgments.

Table 2.2
Means and standard deviations of ratings for forms 2 and 2R of participants' level of agreement with action and inaction in each of the dilemmas as a function of the order in which the dilemma appeared

Dilemma	Order	Action/Inaction	Mean	SD
Trolley	First	Action	3.1	2.6
	Third	Action	1.0	2.9
	First	Inaction	–1.9	2.7
	Third	Inaction	–1.1	3.1
Person	First	Action	–8.6	3.4
	Third	Action	–1.7	4.1
	First	Inaction	–.10	3.5
	Third	Inaction	0.0	3.6
Button[a]	Trolley	Action	2.7	2.8
	Person	Action	.65	3.3
	Trolley	Inaction	–.65	3.3
	Person	Inaction	–2.0	2.8

Positive values indicate agreement, and negative values indicate disagreement.
[a]For the Button dilemma, Order refers to the preceding Dilemma.
Source: Petrinovich & O'Neill, 1994, p. 157.

However, Petrinovich and O'Neill's third pair of forms suggests more subtle analysis. Forms 3 and 3R presented five heterogeneous moral problems (boat, trolley, shield, shoot, shark) in reverse order. Participants' responses to action and inaction in the outside dilemmas did not vary with order. Nonetheless, in the middle shield dilemma, "participants approved of action more strongly (2.6) when it was preceded by the Boat and Trolley dilemmas than when it was preceded by the Shoot and Shark dilemmas (1.0)" (Petrinovich & O'Neill, 1996, p. 160). Some significant framing effects, thus, occur even in heterogeneous sets of moral dilemmas.

In any case, the order of presentation of moral dilemmas does affect many people's moral judgments at least within homogeneous sets of moral problems. Of course, the truth or falsity of moral judgments about actions and inactions in those dilemmas does not depend on which dilemmas preceded or followed the dilemmas in question. Thus, framing effects show ways in which our moral intuitions do not reliably track the truth.

Haidt and Baron

Two more experiments by the Jonathans (Haidt and Baron) also found framing effects in yet another kind of situation. Their first case did not

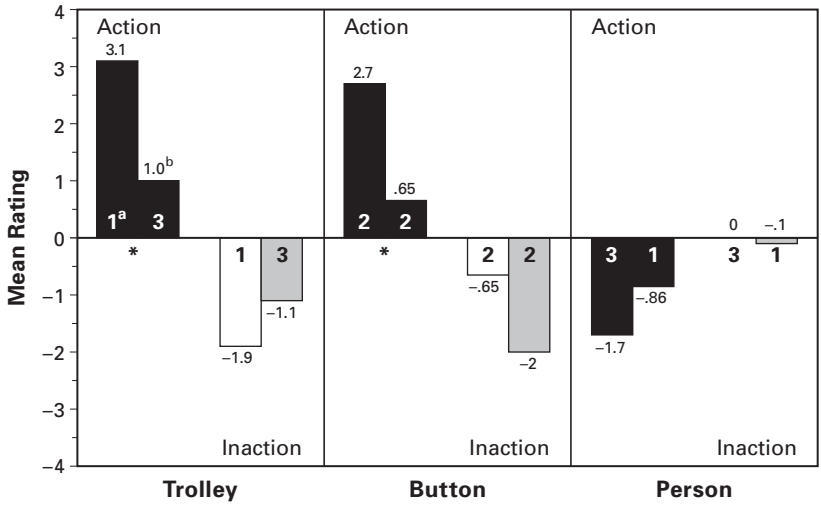

Figure 2.1
Mean ratings for each question for Form 2 and 2R for the Action and Inaction choices in each dilemma. [a]indicates the Order in which the Dilemma appeared in the sequence of questions (1 = first dilemma posed, 2 = second dilemma posed, and 3 = third dilemma posed). [b]indicates the mean rating (positive values indicate agreement with the option, and negative values indicate disagreement). *indicates that the two means differed significantly ($p < .05$). (Reprinted from Petrinovich & O'Neill, 1996, p. 158)

involve killing but only lying. It is also more realistic than most of the other cases in such experiments:

Nick is moving to Australia in two weeks, so he needs to sell his 1984 Mazda MPV. The car has only 40,000 miles on it, but Nick knows that 1984 was a bad year for the MPV. Due to a manufacturing defect particular to that year, many of the MPV engines fall apart at about 50,000 miles. Nevertheless, Nick has decided to ask for $5000, on the grounds that only one-third of the 1984 MPV's are defective. The odds are two out of three that his car will be reliable, in which case it would certainly be worth $5000.

Kathy, one of Nick's best friends, has come over to see the car. Kathy says to Nick: "I thought I read something about one year of the MPV being defective. Which year was that?" Nick gets a little nervous, for he had been hoping that she wouldn't ask. Nick is usually an honest person, but he knows that if he tells the truth, he will blow the deal, and he really needs the money to pay for his trip to Australia. He thinks for a moment about whether or not to tell the truth. Finally, Nick says, "That was 1983. By 1984 they got it all straightened out." Kathy believes him. She likes

the car, and they close the deal for $4700. Nick leaves the country and never finds out whether or not his car was defective. (Haidt & Baron, 1996, pp. 205–206)

Some of the subjects received a different ending:

Nick is trying to decide whether or not to respond truthfully to Kathy's question, but before he can decide, Kathy says, "Oh, never mind, that was 1983. I remember now. By 1984, they got it all straightened out." Nick does not correct her, and they close the deal as before. (Haidt & Baron, 1996, p. 206)

The difference is that Nick actively lies in the first ending whereas he merely withholds information in the second ending. The first version is, therefore, called the act version, and the second is called the omission version.

The relation between Kathy and Nick was also manipulated. In the personal version (as above), Kathy and Nick are best friends. In the intermediate version, Kathy is only "a woman Nick knows from the neighborhood." In the anonymous version, Kathy just "saw Nick's ad in the newspaper." Each of these role versions were divided into act and omission versions.

The six resulting stories were distributed to 91 students who were asked to rate Nick's "goodness" from +100 (maximally good) to 0 (morally neutral) to –100 (maximally immoral). Each subject answered this question about both an act version and an omission version of one of the role variations. Half of the subjects received the act version first. The other half got the omission version first.

The subject's responses are summarized in table 2.3 (from Haidt & Baron, 1996, p. 207). Thus, subjects judged Nick more harshly when he lied than when he withheld information, but the distinction became less important when Nick was good friends with Kathy. They also tended to judge Nick more harshly (for lying or withholding) when he was good friends with Kathy than when they were mere neighbors or strangers. None of this is surprising.

What is surprising is an order effect: "Eighty per cent of subjects in the omission-first condition rated the act worse than the omission, while only 50 per cent of subjects in the act-first condition made such a distinction" (Haidt & Baron, 1996, p. 210). This order effect had not been predicted by Haidt and Baron, so they designed another experiment to check it more carefully.

In their second experiment, Haidt and Baron varied roles within subjects rather than between subjects. Half of the subjects were asked about the act and omission versions with Kathy and Nick as strangers, then about the

Table 2.3
Mean ratings, and percentage of subjects who rated act or omission worse, experiment 1

	Solidarity			
	Anonymous	Intermediate	Personal	Whole Sample
N	31	27	33	91
Act	–53.8	–56.9	–66.3	–59.3
Omission	–27.4	–37.2	–50.8	–38.8
Delta	26.4	19.7	15.5	20.5
Act-worse	74%	67%	52%	64%
Omit-wose	0%	0%	3%	1%

Source: Haidt & Baron, 1994, p. 207.

act and omission versions with Kathy and Nick as casual acquaintances, and finally about the act and omission versions with Kathy and Nick as close friends. The other half of the subjects were asked these three pairs in the reverse order: friends, then acquaintances, and finally strangers.[8] Within each group, half were asked to rate the act first, and the others were asked to rate the omission first.

Haidt and Baron also added a second story that involved injury (but not death or lying). The protagonists are two construction workers, Jack and Ted. The action begins as Ted is operating a crane to move a load of bricks. Here is how the omission version ends:

> Jack is sitting 30 yards away from the crane eating his lunch. He is watching Ted move the bricks, and he thinks to himself: "This looks dangerous. I am not sure if the crane can make it all the way. Should I tell him to stop?" But then he thinks "No, why bother? He probably knows what he is doing." Jack continues to eat his lunch. A few yards short of its destination, the main arm of the crane collapses, and the crane falls over. One of Ted's legs is broken.

Here is the act version:

> Jack is standing 30 yards away from the crane, helping Ted by calling out signals to guide the bricks to their destination. Jack thinks to himself: "[same thoughts]." Jack motions to Ted to continue on the same course [same ending]. (Haidt & Baron, 1996, pp. 208–209)

Haidt and Baron also manipulated the relation between Jack and Ted. Half of the subjects were asked about the act and omission versions with Jack as Ted's boss (the authority version), then about the act and omission versions with Jack as Ted's coworker (the equal version), and finally about the

act and omission versions with Jack as Ted's employee (the subordinate version). The other half of the subjects were asked these three pairs in the reverse order: subordinate, then equal, and finally authority. Within each group, half were asked to rate the act first, and the others were asked to rate the omission first.

The subjects were 48 + 21 students. Because positive ratings were not needed, the scale was truncated to 0 (morally neutral, neither good nor bad) to –100 (the most immoral thing a person could ever do).

The results are summarized in tables 2.4 and 2.5 (from Haidt & Baron, 1996, p. 210). This experiment replicates the unsurprising results from Experiment 1.

More importantly for our purposes, a systematic order effect was found again: "a general tendency for subjects to make later ratings more severe than earlier ratings." This effect was found, first, in the role variations: "In the Mazda story, 88 per cent of subjects lowered their ratings as Nick changed from stranger to friend, yet only 66 percent of subjects raised their

Table 2.4
Mean ratings, and percentage of subjects who rated act or omission worse, experiment 2, Mazda story ($N = 67$)

	Anonymous	Solidarity Intermediate	Personal
Act	–49.2	–54.9	–63.1
Omission	–40.3	–46.9	–57.3
Delta	9.0	7.9	5.9
Act-worse	58%	57%	43%
Omit-worse	2%	0%	0%

Source: Haidt & Baron, 1994, p. 210.

Table 2.5
Mean ratings, and percentage of subjects who rated act or omission worse, Experiment 2, Crane story ($N = 68$)

	Subordinate	Hierarchy Equal	Authority
Act	–41.2	–42.4	–51.9
Omission	–30.4	–31.8	–44.4
Delta	10.8	10.6	7.5
Act-worse	52%	53%	43%
Omit-worse	3%	3%	4%

Source: Haidt & Baron, 1994, p. 210.

ratings as Nick changed from friend to stranger." Similarly, "In the Crane story, 78 per cent of those who first rated Jack as a subordinate lowered their ratings when Jack became the foreman, while only 56 percent of those who first rated Jack as the foreman raised their ratings when he became a subordinate." The same pattern recurs in comparisons between act and omission versions: "In the Crane story, 66 per cent of subjects in the omission-first condition gave the act a lower rating in at least one version of the story, while only 39 per cent of subjects in the act-first condition made such a distinction." In both kinds of comparisons, then, "subjects show a general bias towards increasing blame" (Haidt & Baron, 1996, p. 211).

These changes in moral belief cannot be due to changes in the facts of the case, because consequences, knowledge, intention, and other facts held constant. The descriptions of the cases were admittedly incomplete, so subjects might have filled in gaps in different ways (Kuhn, 1997). However, even if that explains how order affected their moral judgments, order still *did* affect their moral judgments. The truth about what is morally right or wrong in the cases did not vary with order. Hence, moral beliefs fail to track the truth and are unreliable insofar as they are subject to such order effects.

Together these studies show that moral intuitions are subject to framing effects in many circumstances. That is the third premise of the master argument.[9]

The Final Premise

Only one premise remains to be supported. It claims that we ought to know that moral intuitions are subject to framing effects in many circumstances. Of course, those who have not been exposed to the research might not know this fact about moral intuitions. However, this psychological research—like much psychological research—gives more detailed arguments for a claim that educated people ought to have known anyway. Anyone who has been exposed to moral disagreements and to the ways in which people argue for their moral positions has had experiences that, if considered carefully, would support the premise that moral intuitions are subject to framing effects in many circumstances. Those people ought to know this.

Maybe children and isolated or uneducated adults have not had enough experiences to support the third premise of the master argument, which claims that moral framing effects are common. If so, then this argument cannot be used to show that *they* are not justified noninferentially in

trusting their moral intuitions. However, if these were the only exceptions, moral intuitionists would be in an untenable position. They would be claiming that the only people who are noninferentially justified in trusting their moral intuitions are people who do not know much, and they are justified in this way only because they are ignorant of relevant facts. If they knew more, then they would cease to be justified noninferentially. To present such people as epistemic ideals—by calling them "justified" when others are not—is at least problematic. If it takes ignorance to be justified noninferentially, then it is not clear why (or how) the rest of us should aspire to being justified noninferentially.

In any case, if you have read this far, you personally know some of the psychological studies that support the third premise in the master argument. So do moral intuitionists who have read this far. Thus, both they and you ought to know that moral intuitions are subject to framing effects in many circumstances. The last premise and the master argument, therefore, apply to them and to you. They and you cannot be justified noninferentially in trusting moral intuitions. That is what the master argument was most concerned to show.

Responses

Like all philosophical arguments, the master argument is subject to various responses. Some responses raises empirical issues regarding the evidence for moral framing effects. Others question the philosophical implications of those studies.

Psychologists are likely to object that I cited only a small number of studies that have to be replicated with many more subjects and different moral problems. Additional studies are needed not only to increase confidence but also to understand what causes moral framing effects and what does not. Of course, all of the reported results are statistically significant. Moreover, the studies on moral judgments and choices fit well with a larger body of research on framing effects on decisions and judgments in other areas, especially medical and economic decisions (surveyed in Kühberger, 1998, and Kühberger, Schulte-Mecklenbeck, & Perner, 1999). Therefore, I doubt that future research will undermine my premise that many moral beliefs are subject to framing effects. Nonetheless, I am happy to concede that more research on moral framing effects is needed to support the claim that moral beliefs are subject to framing effects in the ways that these initial studies suggest. I encourage everyone (psychologists *and* philosophers) to start doing the research. In the meantime, the trend of the

research so far is clear and not implausible. Hence, at present we have an adequate reason to accept, at least provisionally, the premise that many moral beliefs are subject to framing effects.

More specifically, critics might object that moral believers might not be subject to framing effects when scenarios are fully described. Even if subjects' moral intuitions are not reliable when subjects receive only one description—such as killing or saving—their moral intuitions still might be reliable when they receive both descriptions, so they assess the scenarios within both frames. Most intuitionists, after all, say that we should look at a moral problem from various perspectives before forming a moral judgment. This objection is, however, undermined by Haidt and Baron's second study. Because of its within-subjects design, subjects in that study did receive both descriptions, yet they were still subject to statistically significant framing effects. Admittedly, the descriptions were not given within a single question, but the questions were right next to each other on the page and were repeated in each scenario, so subjects presumably framed the scenarios in both ways. Moreover, in a meta-analysis, Kühberger (1998, p. 36) found "stronger framing effects in the less-frequently used within-subjects comparisons." It seems overly optimistic, then, to assume that adding frames will get rid of framing effects.[10]

The scenarios are still underdescribed in various ways. Every scenario description has to be short enough to fit in an experiment, so many possibly relevant facts always have to be left out. These omissions might seem to account for framing effects, so critics might speculate that framing effects would be reduced or disappear if more complete descriptions were provided. Indeed, Kühberger (1995) did not find any framing effects of wording in the questions when certain problems were fully described. A possible explanation is that different words in the questions lead subjects to fill in gaps in the scenario descriptions in different ways. Kuhn (1997) found, for example, that words in questions led subjects to change their estimates of unspecified probabilities in medical and economic scenarios. If probability estimates are also affected by words and order in moral scenarios, this might explain how such framing affects moral judgments, and these effects would be reasonable if the changes in probability estimates are great enough to justify different moral judgments. Nonetheless, even if this is the process by which framing effects arise, moral intuitions would still be unreliable. Wording and context would still lead to conflicting moral judgments about a single description of a scenario. Thus, it is not clear that this response undermines the master argument, even if the necessary empirical claims do hold up to scrutiny.

Another response emphasizes that the studies do not show that everyone is affected by framing. Framing effects are not like visual illusions that are shared by everyone with normal vision. In within-subjects studies, there are always some subjects who maintain steady moral beliefs without being affected by frames.

But who are they? They might be the subjects who thought more about the problems. Many subjects do not think carefully about scenarios in experimental conditions. They just want to get it over with, and they do not have much at stake. Some moral intuitionists, however, require careful reflection before forming the moral intuitions that are supposed to be justified noninferentially. If moral intuitions that are formed after such careful reflection are not subject to framing effects, then moral intuitionists might claim that the master argument does not apply to the moral intuitions that they claim to be justified noninferentially. In support of this contention, some studies have found that framing effects are reduced, though not eliminated, when subjects are asked to provide a rationale (Fagley & Miller, 1990) or take more time to think about the cases (Takemura, 1994) or have a greater need for cognition (Smith & Levin, 1996) or prefer a rational thinking style (McElroy & Seta, 2003). In contrast, a large recent study (LeBoeuf & Shafir, 2003) concludes, "More thought, as indexed here [by need for cognition], does not reduce the proclivity to be framed" (p. 77). Another recent study (Shiloh, Salton, & Sharabi, 2002) found that subjects who combined rational and intuitive thinking styles were among those *most* prone to framing effects. Thus, it is far from clear that framing effects will be eliminated by the kind of reflection that some moral intuitionists require.

Moreover, if analytic, systematic, or rational thinking styles do reduce framing effects, this cannot help to defend moral intuitionism, because subjects with such thinking styles are precisely the ones who are able to form inferences to justify their moral beliefs. The believers who form their beliefs without inference and those who claim to be justified noninferentially are still subject to framing effects before they engage in such reasoning. That hardly supports the claim that any moral belief is justified noninferentially. To the contrary, it suggests that inference is needed to correct for framing effects. Thus, these results do not undermine the master argument. They support it.

Finally, suppose we do figure out which people are not subject to moral framing effects. Moral intuitionism still faces a dilemma: if we can tell that we are in the group whose moral intuitions are reliable, then we can get inferential confirmation; if we cannot tell whether we are in the group

whose moral intuitions are reliable, then we are not justified. Either way, we cannot be justified independent of inferential confirmation.

To see the point, imagine that you have a hundred old thermometers.[11] You know that many of them are inaccurate, though you don't know exactly how many. It might be eighty or fifty or ten. You pick one at random, put it in a tub of water, which you have not felt. The thermometer reads 90°. Nothing about this thermometer in particular gives you any reason to doubt its accuracy. You feel lucky, so you become confident that the water is 90°. Are you justified? No. Since you believe that a significant number of the thermometers are unreliable, you are not justified in trusting the one that you happen to randomly pick. You need to check it. One way to check it would be to feel the water or to calibrate this thermometer against another thermometer that you have more reason to trust. Such methods might provide confirmation, and then your belief might be justified, but you cannot be justified without some kind of confirmation.

In addition to having confirmation, you need to know that it is confirmation. To see why, imagine that the thermometers are color coded. Their tops are red, orange, yellow, green, and blue. All of the blue and green thermometers are accurate, some but not all of the yellow ones are accurate, but none of the red ones are accurate. However, you are completely unaware of any relation between colors and accuracy. Then you randomly pick a blue one, see its top, and trust it. Even though you know it is blue, and its being blue would give you good evidence that it is accurate if you knew that all the blue thermometers are accurate, still, if you do not know that its being blue is good evidence of its accuracy, then you are unjustified in trusting this thermometer. Thus, it is not enough to have a belief that supports accuracy. You need to know that it supports accuracy.

These thermometers are analogous to the processes by which believers form immediate moral beliefs. According to moral intuitionism, some moral believers are justified in forming immediate moral beliefs on the basis of something like (though not exactly like) a personal moral thermometer that reliably detects moral wrongness and rightness. However, the analogy to the hundred thermometers shows that, if we know that a large number of our moral thermometers are broken or unreliable in many situations, then we are not justified in trusting a particular moral thermometer without confirmation. Maybe we got lucky and our personal moral thermometer is one of the ones that works fine, but we are still not justified in trusting it, if we know that lots of moral thermometers do not work, and we have no way of confirming which ones do work. This standard applies to moral beliefs, because we do know that lots of moral

thermometers do not work. That's what framing effects show: our moral beliefs must be unreliable when they vary with wording and context. The range of framing effects among immediate moral beliefs thus shows that many of our moral thermometers are unreliable. It doesn't matter that we do not know exactly how many are unreliable or whether any particular believer is unreliable. The fact that moral framing effects are so widespread still reveals enough unreliability to create a need for confirmation of moral beliefs, contrary to moral intuitionism.

Critics might complain that, if my own moral intuition is reliable and not distorted, then I am justified in trusting it, because it is mine. But recall the colored thermometers. Merely knowing a feature that is correlated with accuracy is not enough to make me justified. I also need to know that this feature is correlated with accuracy. The same standard applies if the feature that is correlated with accuracy is being my own intuition. In the moral case, then, I need to know that my moral intuition is reliable. If I know that, then I have all the information I need in order to make me able to justify my belief with an inference. Thus, I am not justified noninferentially in trusting my own moral intuition.

This point also applies to those who respond that some moral intuitions are not subject to framing effects. All that moral intuitionists claim is that some moral intuitions are reliable. The studies of framing effects show that some moral intuitions are not reliable. Maybe some are and others are not. Thus, the studies cannot refute the philosophical claim. More specifically, the studies suggest *which* moral intuitions are not subject to framing effects. Recall the transplant case in Petrinovich and O'Neill's nonhomogeneous Forms 1 and 1R. They found no framing effects there—so maybe moral intuitions like these are justified noninferentially, even if many others are not.

This response runs into the same dilemma as above: if a particular moral intuition is in a group that is reliable or based on a reliable process, then the person who has that moral intuition either is or is not justified in believing that it is in the reliable group. If that person is not justified in believing that it is in the reliable group, then he is not justified in trusting it. However, if he is justified in believing that this moral intuition is in the reliable group, then he is able to justify it by an inference from this other belief. Either way, the moral believer is not justified independent of inferential confirmation. That is all that the master argument claims.

This argument might not seem to apply to moral intuitionists who claim only that general prima facie (or pro tanto) moral principles can be justified noninferentially. Standard examples include "It is prima facie morally

wrong to kill" and "It is prima facie morally wrong to lie." If such moral principles are justified by intuitive induction from specific cases, as Ross (1939, p. 170) claimed, then they will be just as unreliable as the specific cases from which they are induced. However, if moral intuitions of general principles are supposed to be justified directly without any reference at all to specific cases, then the above experiments might seem irrelevant, because those experiments employ particular cases rather than general principles. This response, however, runs into two problems. First, such general principles cannot be applied to concrete situations without framing the information about those situations. What counts as killing depends on the baseline, as we saw. However, if such general principles cannot be applied without framing effects, then it seems less important whether their abstract formulations are subject to framing effects. In any case, even though current studies focus on concrete examples rather than general principles, general principles could be subject to framing effects as well. They are also moral intuitions after all. Hence, since many other moral intuitions are subject to framing effects, it seems reasonable to suppose that these are, too, unless we have some special reason to believe that they are exempt. But if we do have a special reason to exempt them, then that reason makes us able to infer them in some way—so we arrive back at the same old dilemma in the end.

Finally, some moral intuitionists might accuse me of forgetting that believers can be *defeasibly* justified without being *adequately* justified. A believer is defeasibly justified whenever the following counterfactual is true: the believer would be adequately justified if there were no defeater. If moral believers would be adequately justified in the absence of any framing effect, then, even if framing effects actually keep moral believers from being adequately justified apart from inferential confirmation, those moral believers still might be defeasibly justified apart from inferential confirmation.

However, it is crucial to distinguish two kinds of defeaters. An *overriding* defeater of a belief provides a reason to believe the opposite. In contrast, an *undermining* defeater takes the force out of a reason without providing any reason to believe the opposite. For example, my reason to trust a newspaper's prediction of rain is undermined but not overridden by my discovery that the newspaper bases its prediction on a crystal ball. This discovery leaves me with no reason at all to believe that it will rain or that it will not rain. Similarly, the fact that moral intuitions are subject to framing effects cannot be an overriding defeater, because it does not provide any reason to believe that those moral intuitions are false. Thus, framing effects

must be undermining defeaters. But then, like the discovery of about the crystal ball, moral framing effects seem to leave us with no reason to trust our immediate moral beliefs before confirmation.

Moral intuitionists can still say that some immediate moral beliefs are defeasibly justified if that means only that they *would* be adequately justified *if* they were not undermined by the evidence of framing effects. This counterfactual claim is compatible with their actually not being justified at all, but only appearing to be justified. Such moral believers might have no real reason at all for belief but only the misleading appearance of a reason, as with the newspaper's weather prediction based on a crystal ball. That claim is too weak to worry about.

Besides, even if we did have some reason to trust our moral intuitions apart from any inferential ability, this would not make them adequately justified. Skeptics win if no moral belief is adequately justified. Hence, moral intuitionists cannot rest easy with the claim that moral intuitions are merely defeasibly justified apart from inferential ability.

Conclusions

I am not claiming that no moral beliefs or intuitions are justified. That academic kind of moral skepticism does not follow from what I have said here. Moreover, I do not want to defend it. My point here is not about *whether* moral beliefs are justified but rather about *how* they can be justified. I have not denied that moral beliefs can be justified inferentially. Hence, I have not denied that they can be justified.

What I am claiming is that no moral intuitions are justified noninferentially. That is enough to show why moral intuitionism (as I defined it) is false. Moral intuitionists claim that moral intuitions are justified in a special way: without depending on any ability to infer the moral belief from any other belief. I deny that any belief is justified in that way.

Behind my argument lies another claim about methodology. I am also claiming that empirical psychology has important implications for moral epistemology, which includes the study of whether, when, and how moral beliefs can be justified. When beliefs are justified depends on when they are reliable or when believers have reasons to believe that they are reliable. In circumstances where beliefs are based on processes that are neither reliable nor justifiably believed to be reliable, they are not justified. Psychological research, including research into framing effects, can give us reason to doubt the reliability of certain kinds of beliefs in certain circumstances. Such empirical research can, then, show that certain moral beliefs are not

justified. Moral intuitionists cannot simply dismiss empirical psychology as irrelevant to their enterprise. They need to find out whether the empirical presuppositions of their normative views are accurate. They cannot do that without learning more about psychology and especially about how our moral beliefs are actually formed.

Notes

1. For a systematic critique of attempts to justify moral theories without appealing to moral intuitions, see Sinnott-Armstrong (2006).

2. Some defenders of moral intuitions do not count anything as a moral intuition unless it is true or probable or justified. Such accounts create confusion when we want to ask whether moral intuitions are reliable or justified, because an affirmative answer is guaranteed by definition, but skeptics can still ask whether any people ever have any "real" moral intuitions. To avoid such double-talk, it is better to define moral intuitions neutrally so that calling something a moral intuition does not entail by definition that it has any particular epistemic status, such as being true or probable or justified.

3. Contrary to common philosophical dogma, there is a logically valid way to derive a moral "ought" from "is," but such derivations still cannot make anyone justified in believing their conclusions. See Sinnott-Armstrong (2000).

4. Van Roojen might admit that Horowitz's argument undermines moral intuitionism, since he defends a method of reflective equilibrium that is coherentist rather than foundationalist.

5. Another possible explanation is change in beliefs about probabilities. See Kuhn (1997). However, this would not cover all of the moral cases and would not save the reliability of moral intuitions anyway.

6. Kamm gives many other examples and arguments, but I cannot do justice to her article here. For further criticisms, see Levy (forthcoming).

7. To disagree with an alternative is, presumably, to see it as morally wrong. However, this is not clear, since subjects were asked what they would do—not what was wrong.

8. To make it clearer that Nick would not have told the truth if Kathy had not interrupted, the omission version was changed to read, ". . . Nick decides to lie to Kathy, but [before Nick can speak] Kathy says, 'Oh, never mind, that was 1983.' "

9. Unger (1996) argues that many other moral intuitions change when intervening cases are presented between extremes. If so, these cases present more evidence of framing effects. A final bit of evidence for framing effects comes from philosophical paradoxes, such as the mere addition paradox (Parfit, 1984). In Parfit's example,

when people compare A and B alone, most of them evaluate A as better. In contrast, when people consider B+ and A– in between A and B, most of them do not evaluate A as better than B. The fact that Parfit's paradox still seems paradoxical to many philosophers after long reflection shows how strong such framing effects are.

10. For more on framing effects when both frames are presented, see Armstrong, Schwartz, Fitzgerald, Putt, and Ubel (2002), Druckman (2001), and Kühberger (1995).

11. My analogy to thermometers derives from Goldman (1986, p. 45). The same point could be made in terms of fake barns, as in Goldman (1976).