American Dialectic, Vol. 4, No.1, 2014 (1-27)

A Kantian Take on Fallible Principles and Fallible Judgments

Samuel Kahn

In the second half of the *Metaphysics of Morals*, Kant makes the following claim about acting in accordance with conscience:

> But if someone is aware that he has acted in accordance with his conscience, then as far as guilt or innocence is concerned nothing more can be required of him. (6:401)[1]

This claim is striking given that immediately before making it, Kant admits that it is possible for an agent to believe that some action X is right even though it is an objective truth that X is not right: "I can indeed be mistaken at times in my objective judgment as to whether something is a duty or not" (6:401).

According to Kant, agents do not have infallible knowledge of right and wrong, but if they act in accordance with conscience, then they have done all that they ought as far as morality is concerned. One can understand this more clearly by means of the distinction between objective and subjective rightness.[2]

Philosophers often distinguish between two different senses of "rightness": "objective" and "subjective." Suppose some action D is not objectively right. Suppose, further, that it is possible for me to believe that D is right and that I do believe that it is so. If in these

[1] Compare, for example, 6:189: "more [than acting in accordance with conscience] cannot be required of a human being." See also 27:335 and 355. Throughout this paper, references to Kant's works cite volume and page number of the standard Academy edition. Translations are taken from The Cambridge Editions of the Works of Immanuel Kant (trans. Paul Guyer and Allen W. Wood).

[2] Kantians might substitute 'permissibility' for 'rightness' to avoid ambiguity; I shall not be talking about Kant's duties of right, which are more properly discussed in the context of his *Rechtslehre*.

conditions I perform D, then I have acted subjectively but not objectively rightly.[3] There are two substantive issues here: whether agents can be mistaken about objective rightness (whether there is such a thing as subjective rightness) and what to say about agents who make such mistakes. In the passages quoted above, Kant is saying that (1) objective rightness and subjective rightness sometimes come apart and (2) if an agent acts in accordance with his conscience (and, thus, with that which he judges to be right) then he has done all that he ought as far as morality is concerned.

As is hopefully now clear, taken together, these two positions entail that it might be the case for some agent A, A can perform some objectively wrong action X blamelessly. Lest there be any doubt that Kant would accept this, it should be noted that Kant articulates both of these positions throughout his corpus and, moreover, in the passage excerpted above, he does so in such close succession as to make it almost unthinkable that he did not realize what they entail.[4]

In this paper I explore Kant's doctrine more fully in order to determine whether it is defensible. In particular, I confront two issues: the blameworthiness of acting contrary to fallible knowledge and the blamelessness of acting according to a fallible judgment.[5]

---

[3] For a good discussion of the distinction between subjective and objective rightness (independent of Kant), see Mark van Roojen, "Moral Rationalism and Rational Amoralism," *Ethics* 120 (2010): 495-525.

[4] For a more thorough discussion of the textual issues here, see Samuel Kahn, "Kant's Theory of Conscience" in *Rethinking Kant: Volume IV*, ed. Pablo Muchnik and Oliver Thorndike (Cambridge Scholars Publishing, Forthcoming).

[5] There is some common ground between the issues I confront in this paper and recent debates on Kantian moral worth. However, I shall not be engaging with any of these debates and I shall not be discussing Kantian moral worth as such. For some of the more important, recent contributions to these debates, see, for example, Marcia Baron, *Kantian Ethics Almost without Apology* (Ithaca: Cornell University Press, 1995), chapter 1; Richard Henson, "What Kant Might Have Said: Moral Worth and the Overdetermination of Dutiful Actions," *The Philosophical Review* 88, no. 1 (1979): 39-54; Barbara Herman, *The Practice of Moral Judgment* (Cambridge, Massachusetts: Harvard University Press, 1993), chapter

I. Acting contrary to fallible knowledge

As pointed out above, according to Kant if an agent acts according to his conscience, then he has done all that he ought as far as morality is concerned. Moreover, Kant holds that this is so despite the fact that an agent who acts according to conscience might perform an action that is objectively wrong. Now strictly speaking, nothing that Kant says in his discussion of conscience commits him to the claim that if an agent does not act according to his conscience, then he has not done all that he ought as far as morality is concerned.[6] But it is difficult to see how Kant could avoid committing to this claim. Any plausible defense of Kant's views on conscience surely would make appeal to the notion of autonomy, and given the stark contrast Kant makes between autonomy and heteronomy in the *Groundwork for a Metaphysics of Morals*, it is difficult to see how he could pull back from commitment to the converse of his conscience claim, and this, of course, would commit him to the inverse.

However, this might seem counterintuitive in cases in which agents perform objectively right actions precisely because they are not acting in accordance with their consciences. As soon as subjective rightness and objective rightness have been pried apart, one must allow for the possibility of an agent acting subjectively wrongly but objectively rightly. The worry is that pretheoretic intuition rests in favor of saying that (at least some) such agents are

1; and Allen Wood, *Kant's Ethical Thought* (Cambridge: Cambridge University Press, 1999), chapter 1.

[6] I am aware of only a single place in Kant's corpus where he says something along these lines, and it is in the *Lectures on Ethics* rather than in Kant's published works. The passage runs as follows: "if I am convinced, for example, in my conscience, that to prostrate oneself before images is idolatry, and I am in a place where this is going on, then if I did it in order to give no offense to others, I would be acting against my conscience; *yet this must be holy to me*. I can, indeed, feel sorry that any one should be offended thereby, but it is no fault of mine" (27:335, my emphasis).

praiseworthy whereas Kant's position requires saying that (all) such agents are blameworthy.

Since Bennett's seminal article on conscience, Huck Finn, who resolves to act subjectively rightly (and hence objectively wrongly) but acts akratically (and hence objectively rightly), has been taken as the paradigm example of an agent who acts in a praiseworthy way despite disobeying the voice of his conscience.[7] To recall the story, Huck Finn has been helping his friend Jim, a runaway slave,[8] to escape to the North. They are journeying by raft down the Mississippi river and (in chapter 16) they are nearing freedom when Huck begins to feel conflicted. He wonders whether he has done the right thing in helping Jim to escape and he begins to think that he ought to turn Jim in. Huck tells us, "my conscience got to stirring me up hotter than ever, until at last I says to it: 'Let up on me—it ain't too late, yet—I'll paddle ashore at first light, and tell.' I felt easy, and happy, and light as a feather, right off. All my troubles was gone."

The episode in which Huck tries to make good on this resolve is the one that has attracted so much attention since Bennett. As the story goes, Huck sets off to turn Jim in at first light, but at the critical moment Huck's resolve weakens in the face of sympathy for his friend. Thus, despite having judged it to be subjectively right to turn Jim in and having resolved to act accordingly, Huck does not turn Jim in and, indeed, continues to help Jim to escape to freedom.

---

[7] See Jonathan Bennett, "The Conscience of Huckleberry Finn," *Philosophy* 49, no. 188 (1974): 123-34. For some of the early reactions to this article, see, for example, John Harris, "Principles, Sympathy and Doing What's Right," *Philosophy* 52, no. 199 (1977): 96-99 and Jenny Teichman, "Mr. Bennett on Huckleberry Finn," *Philosophy* 50, no. 193 (1975): 358-59.

[8] The novel is somewhat complicated on this front. Both Huck and Jim believe that Jim is a runaway slave. But later in the story it turns out that Jim was free all along. But the whole point is that Huck does not know this.

The basic issue is that Huck seems to be speaking honestly and ingenuously in saying that his conscience told him to turn Jim in and that he feels guilty for not doing so; it seems to follow that anyone (such as Kant) who believes that people are required to follow their consciences must be committed to saying that Huck's conduct in not turning Jim in is blameworthy. But neither Mark Twain nor any morally sensitive reader of Mark Twain's novel is likely to want to say this; most think it was admirable of Huck to ignore his conscience and to help Jim gain his freedom and that it would have been downright despicable of Huck to act as his conscience told him. Readers approve of Huck because he was unable (by his own account, morally too weak) to follow his conscience.

There are of course many different interpretations of this episode (and of Huck Finn's character in general) on hand in the secondary literature.[9] The interpretation I just went through is

[9] For a small cross-section, see Nomy Arpaly and Timothy Schroeder, "Praise, Blame and the Whole Self," *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition* 93, no. 2 (1999): 161-88; Nomy Arpaly, "On Acting Rationally against One's Best Judgment," *Ethics* 110, no. 3 (2000): 488-513; Nomy Arpaly, "Moral Worth," *The Journal of Philosophy* 99, no. 5 (2002): 223-45; Ronald de Sousa and Adam Morton, "Emotional Truth," *Proceedings of the Aristotelian Society*, Supplementary Volumes, 76 (2002): 247-63 & 265-75; Mathieu Doucet, "In Praise of Akrasia?" Unpublished manuscript; Carol Freedman, "The Morality of Huck Finn," *Philosophy and Literature* 21, no. 1 (1997): 102-13; Alan Goldman, "Huckleberry Finn and Moral Motivation," *Philosophy and Literature* 34, no. 1 (2010): 1-16; Chad Kleist, "Huck Finn the Inverse Akratic: Empathy and Justice," *Ethical Theory and Moral Practice* 12 (2009): 257-66; Jung H. Lee, "The Moral Power of Jim: A Mencian Reading of Huckleberry Finn," *Asian Philosophy* 19, no. 2 (2009): 101-118; Philip Montague, "Re-Examining Huck Finn's Conscience," *Philosophy* 55, no. 214 (1980): 542-46; and Craig Taylor, "Moral Incapacity and Huckleberry Finn," *Ratio* (new series) XIV, 1 (2001): 56-67. My interpretation is inspired from the work of Holton (Richard Holton, "Inverse Akrasia and Weakness of Will," Unpublished manuscript). Some authors do make claims about Huck Finn that I think are either directly contradicted or wholly unsupported by the text. For example, Kleist argues that "Huck . . . never believes it is right for him to say "I'm sorry" to Jim" (Kleist, "Huck Finn the Inverse Akratic," 263); Kleist seems to be forgetting the scene in chapter 15 in which Huck does believe it is right to apologize to Jim

Bennett's, and if there is a standard reading of this episode from chapter 16, it is Bennett's. I shall explore this reading in more detail shortly.

Agents like Huck Finn are often referred to as inverse akratics.[10] In labeling Huck and others as inverse akratics, the idea is not that there is something good about their action *qua* akratic. The idea is that in some cases the akratic course of action is superior to the course of action recommended by an agent's best judgment.[11] That is, an inverse akratic acts objectively rightly precisely because he is acting against his resolve to act according to his best judgment;[12] in such cases the action is not good because it is performed akratically but rather despite the fact that it is performed akratically.[13] Huck's case is particularly poignant because on the standard reading his best judgment, that which is subjectively right for him, is clearly objectively wrong; moreover, he is "driven" to act akratically (and, thereby, objectively rightly) by his broad

and, indeed, actually brings himself to do so. But my goal in this paper is to illustrate Kant's theory of conscience; I explore Huck Finn only insofar as I think he can be used to serve this purpose.

[10] I believe Arpaly and Schroeder coined the term 'inverse akrasia' in "Praise, Blame and the Whole Self" (see, especially, 162).

[11] Ibid.

[12] Compare Arpaly, "On Acting Rationally."

[13] In "Praise, Blame and the Whole Self," Arpaly and Schroeder are careful not to say that there is anything intrinsically good or bad about akrasia. Indeed, they remark that inverse akrasia and normal akrasia occur in contexts that have nothing to do with morality (163). This might be taken as evidence that they do not take akrasia to be intrinsically good or bad. Thus, one might wonder whether Arpaly and Schroeder really do think that cases of inverse akrasia are good *despite* being akratic. However, in another article, Arpaly claims, "sometimes, an agent is more rational for acting against her best judgment than she would be if she acted in accordance with her best judgment. I still agree that every agent who acts against her best judgment is, as an agent, less than perfectly rational . . . however, I will argue that it is not always the case that an agent is less rational (and less coherent) in following the desire than she is in following her best judgment" (Arpaly, "On Acting Rationally," 491; see also 492). Thus, when push comes to shove, Arpaly, anyway, really does seem to think that cases of inverse akrasia are good despite being cases of akrasia; there is nothing good, on her account, about akrasia.

sympathies. Huck resolves to act according to his principles, which prescribe turning Jim in to the slave-catchers, but his sympathy and fellow feeling for Jim win the day, and in the end he helps Jim to escape. And because Huck's principles are so inhumane, there is thought to be something particularly morally praiseworthy about his acting in accordance with his sympathies despite the fact that by so doing he is acting against his principles.

Again, the fact that Huck is judged morally praiseworthy for helping Jim to escape is a problem for Kant. It looks like Kant should say that all cases of inverse akrasia are morally blameworthy. But cases like Huck reveal that at least some might be morally praiseworthy. In trying to resolve this problem, I think that there are two things to say. The first has to do with the fictional case of Huck Finn as portrayed by Mark Twain in particular; the second has to do more generally with agents who act subjectively wrongly but objectively rightly as a result of acting against fallible principles.

To return to the story, the morning ("first light") after Huck made his resolve comes around and Huck sets off, telling Jim (falsely) that he is going to make sure that it is safe enough for them to continue their journey. As Huck pushes out with the canoe, Jim thanks him and tells him that he is the best friend "ole Jim" ever has had. Huck recounts his reaction to Jim's words as follows:

> I was paddling off, all in a sweat to tell on him; but when he says this, it seemed to kind of take the tuck all out of me. I went along slow then, and I warn't right down certain whether I was glad I started or whether I warn't. When I was fifty yards off, Jim says: 'Dah you goes, de ole true Huck; de on'y white genlman dat ever kep' his promise to ole Jim.' Well, I just felt sick. But I says, I got to do it—I can't get out of it.[14]

[14] Kleist argues that 'it' refers to "Huck's never-ending compassion to keep Jim safe" (Kleist, "Huck Finn the Inverse Akratic," 259). I disagree with Kleist's reading. It is crucial to understanding Huck's character and the conflict he experiences here that 'it' be understood to refer to the act of turning Jim in to the authorities. Huck is trying to screw himself to the sticking-place, steeling himself to perform a difficult deed.

Despite his efforts to bring himself to turn Jim in, Huck cannot do so in the end. When Huck gets within speaking distance of the two men who are hunting runaway slaves, they ask him whether the man on his raft is white or black, and Huck recounts the experience as follows:

> I didn't answer up prompt. I tried to, but the words wouldn't come. I tried, for a second or two, to brace up and out with it, but I warn't man enough—hadn't the spunk of a rabbit. I see I was weakening; so I just give up trying, and up and says: 'He's white.'

The reading of this passage seems clear enough; Huck's conscience tells him to turn Jim in, but he feels his resolve weakening and he just gives up, goes with the flow of his sympathies. That, anyway, is what seems to be happening—until Huck returns to Jim and makes the following remarks:

> I got aboard the raft, feeling bad and low, because I knowed very well I had done wrong, and I see it warn't no use for me to try to learn to do right . . . then I thought a minute, and says to myself, hold on — s'pose you'd a done right and give Jim up; would you feel better than what you do now? No, says I, I'd feel bad—I'd feel just the same way I do now.

This must give pause. Huck says that his conscience would have plagued him regardless of what he did. If he had turned Jim in, his conscience would have judged him blameworthy for turning Jim in; having helped Jim to escape, his conscience judges him blameworthy for helping Jim escape. Indeed, in a revision of the book that Twain prepared for a lecture circuit, the following words are put into Huck's mouth:

> I don't want no such thing around as a conscience. . . . You ain't wanted, you ain't welcome, you ain't no use to me. I never see such a low-down troublesome cuss, I says. It don't make no difference what a person does, you ain't ever satisfied and you is as free as if you owned the whole layout. If I'd a give Jim up you'd a kep me awake a week mournin' about it; and now you're gittin' ready to try to keep me awake another week because I *didn't* give him up. . . . I wouldn't be as ignorant as you for wages. You

> don't know right from wrong, you ain't got no judgment, you ain't got no sense about anything—you ain't no good but just to lazy around, find fault and keep a person in a sweat.[15]

Moreover, in chapter 33 of the published version of the text, Huck makes the following remark:

> It don't make no difference whether you do right or wrong, a person's conscience ain't got no sense, and just goes for him anyway. If I had a yaller dog that didn't know no more than a person's conscience does, I would pison him.

So far from being a clear-cut case of inverse akrasia, Huck seems not to know whether what he is doing is right or wrong when he sets off in his canoe to turn Jim in.[16] Huck's remarks about conscience reveal that he does not trust it as a good indicator of right and wrong. More to the point, Huck's remarks reveal that he is clearly assuming that conscience *tells* one to do things — that it includes the judgment that X is one's duty. Some conceptions of conscience would allow this, but Kant's does not. That is highly relevant, for it means Huck is mixing up the "telling one to do" (that is, the judgment of understanding about one's duty) with the judicial function of conscience (that makes Huck feel bad).

In order to determine what conclusions Kant's ethical theory requires about Huck, these two things (the judgment of understanding and the judicial function of conscience) must be teased apart, for the question for Kant is whether the inner judicial process that makes Huck think he is blamable for protecting Jim is a complete and legitimate one. One reason for doubting that it is consists in the fact that Huck sees that he also would judge himself

[15] Quoted in Holton, "Inverse Akrasia and Weakness of Will."
[16] In the published text, Huck says, "I went along slow then, and I warn't right down certain whether I was glad I started or whether I warn't." In the revised version of the text for his lecture circuit, Twain amends this passage to read, "It kind of all *unsettled* me, and I couldn't seem to *tell* whether I was doing *right* or doing *wrong*" (quoted in Holton, "Inverse Akrasia and Weakness of Will").

blamable if he had turned Jim over to the slave-catchers. I think Kant would (or ought to) say that this shows that Huck never fully rendered a judgment of conscience at all. His inner judicial process resulted in no more than provisional and conflicting judgments that were never finally resolved. It is as if a judge, at the point where the verdict is to be rendered, summarized the case in two conflicting ways and failed to make a decision. It is essential to Kant's claim that agents who act according to conscience have done all that they ought as far as morality is concerned that conscience must judge (unambiguously). If it does not, then there is nothing that either could be correct (in accordance with conscience) or in error (contrary to conscience). Conflicting judgments are just as indecisive as none at all.

Huck's remarks about conscience thus reveal that he has what Kant might call a diseased conscience.[17] That is, Huck's conscience makes him feel guilty all the time, at least about certain matters, regardless of what he does. If the evidence for reading Huck as an inverse akratic is that he does something praiseworthy despite acting against his resolve to follow his conscience, then this is no evidence at all, for Huck would have been acting against his conscience no matter what he did.

Although this reading puts me at variance with the traditional interpretation of Huck Finn, it is nonetheless revelatory of an important issue for Kant's theory of conscience. The issue is what Kant should say about any agent who has a diseased conscience.

[17] Huck's remarks about poisoning his conscience like a "yaller dog" recall the following claim Kant makes in describing a morbid conscience: "Those who have a tormenting conscience eventually weary of it and finally send it on vacation" (27:357). But the rest of Kant's description shows that by 'morbid conscience' he seems to mean a conscience that finds evil everywhere in one's actions, perhaps by holding one to principles too strict to be reasonable (related to the moral enthusiasm discussed at 6:409). That is not quite what is going on with Huck. It does not seem accurate to say that Huck is too strict with himself.

Kant does not consider the idea that an agent might have a diseased conscience when he claims that an agent who acts in accordance with his conscience has done all that he ought as far as morality is concerned. But if the example of Huck is anything to go by, the trouble arises mainly with regard to the inverse of this claim, for an agent with a diseased conscience always might be at variance with his conscience. Clearly if Huck is taken at his word, then his conscience is not a suitable judge of whether he is blameworthy. Perhaps if an agent has a diseased conscience, then all bets are off; acting in accordance with a diseased conscience does not guarantee blamelessness and acting contrary to a diseased conscience does not guarantee blameworthiness.

The trouble with this conclusion, however, is that conscience seems to be so vitally connected to the notion of autonomy on Kant's account that it seems more proper to say that a diseased conscience compromises the very agent-hood of an agent. This would not be very disturbing in Huck's case; Huck is still quite young. One might say that Huck is too young to be counted a full blown rational agent anyway. Indeed, this point is often overlooked in the secondary literature on Huck Finn. Huck is at most an adolescent, and he is not an educated or mature one. This means that we should not automatically transfer our reactions to him (what we approve or condemn, or feel we should not condemn in his thoughts or conduct) to full-fledged adults. However, this reveals why one might find this conclusion troubling; an adult might be held more responsible than we hold Huck for accepting the erroneous views of the surrounding society. We might wink at a child's acting against his conscience in a way we would not at an adult acting against his conscience. These are important points, for Kant's theory is meant as an account of the conscience of fully adult moral agents, whose responsibility for their actions is more complete than a child's.

Of course, most mature adults, considering the issue, would not follow Huck in coming down flatly on the side of saying that conscience is simply a bad thing to have.[18] But forming that judgment is neither necessary nor sufficient for one to be considered as having a diseased conscience. Reflecting on exactly what is meant here by a diseased conscience might make this potentially troubling conclusion more attractive. An agent with a diseased conscience is unable to reach a conclusion about whether a certain course of action is blameworthy. He simply feels guilty regardless of which course of action he pursues, for his conscience is too indecisive to reach a static conclusion about what is blameworthy. Such an agent really does seem to have something wrong with him; there seems to be nothing to do except to incorporate an exception clause into Kant's theory of conscience. Perhaps when dealing with an agent with a diseased conscience, the agent must be judged simply with regard to that which the agent ought to judge subjectively right. I shall come back to this in the next section.

This is the first point. Although Huck is generally taken as an archetypal case of inverse akrasia, a close examination reveals that he has a diseased conscience and is genuinely unsure about what he ought to do rather than that he is acting objectively rightly but contrary to a resolve to act subjectively rightly. Consideration of Huck Finn in this light (as an agent with a diseased conscience)

[18] There is some evidence that Twain was putting his own thoughts into Huck's mouth with regard to conscience (see Holton, "Inverse Akrasia and Weakness of Will"). But Twain knows that he and his reader, especially if the reader is an adult like himself, will find some of what Huck says amusing and take it ironically, which means Twain is not simply expressing his own thoughts in a form he might have them. Perhaps Twain thinks that he and his adult readers sometimes find themselves in situations where they are likely to blame themselves and have conscience trouble whatever they do, and he wants us to reflect on this situation, perhaps questioning whether in such situations conscience is really a good thing to have. But Huck expresses such doubts about the value of conscience in a way that only a child or immature person could express them, with a kind of innocence that it seems highly unlikely Mark Twain could share or expect his readers to share.

reveals that Kant's theory of conscience probably needs some sort of exception clause for agents with diseased consciences. An agent with a diseased conscience might be judged to be blameworthy by his conscience regardless of what he does; but it is counterintuitive to say that a diseased conscience is correct in its verdict. The point of labeling a diseased conscience as diseased is to bring out the fact that it is abnormal (in a bad way) and, hence, that it should not be taken as indicative of an agent's genuine blameworthiness.

Although this might answer the problem posed by Huck Finn in particular, it leaves unsolved the problem of inverse akrasia in general. There very well might be cases that would illustrate the Huck-Jim situation as it is interpreted (or misinterpreted) by many authors. For instance, there does not seem to be anything obviously absurd about a Huck whose thoughts and actions fit what Bennett (wrongly) says Huck's were. And as moral philosophers, we must have a justified view about what to say about those other cases too. A great work of fiction often brings in (and leaves out) the right things, capturing the psychology of human action better than it is captured by more simple-minded examples motivated only by theories and theory-driven intuitions. But interpreting Huck Finn as I have done, although revelatory for one aspect of Kant's theory of conscience, leaves untouched the issue of inverse akrasia, a real issue that deserves consideration.

For a clearer case of inverse akrasia, consider Heinrich Himmler, Reichsführer of the SS and leading member of the Nazi party. Himmler described himself as suffering from frequent bouts of weakness of will, during which his resolve to act according to his principles was weakened by his natural sympathies for the Jews.[19]

[19] See, for example, Heinrich Fraenkel and Roger Manwell, *Heinrich Himmler* (London: Heinemann, 1965), 132 and 187; or William Shirer, *The Rise and Fall of the Third Reich* (New York: Simon and Schuster, 1960), 966.

Suppose that Himmler is taken at his word; he really does have the atrocious principles that he professes to have. Suppose further that in one of his fits of weakness of will, Himmler aids a Jewish family in getting a passport to leave the country (he did not actually do this, but suppose that he did). After the family escapes, Himmler is tormented by remorse for what he did; he has undercut his own goal of "purifying" the human race. Must Kant say that Himmler should not have helped the family to escape? This is the problem posed by the inverse akratic. Given Himmler's character and (other) actions, one probably would not praise him very highly for his momentary "lapse." But surely one ought not to condemn him for acting against his principles and thereby saving some Jews from the concentration camps.

In fact, I think Kant's theory (without any modification or exception clauses) gives the correct answer here. Himmler has put himself in a situation in which he cannot act rightly no matter what he does. It is a bad thing that Himmler goes against his conscience even though most do not approve of what his conscience tells him and do approve of the action he does in violation of it. It would have been better if Himmler's conscience had told him the right thing (namely, not to massacre the Jews), if this had been the result of his best judgment and if he had done it with the approval of conscience. In fact, this last scenario is the only one where everything about Himmler's action is to be approved. Acting against one's resolution to act subjectively rightly is always a bad thing even when it leads to an action that happens to be good, for in such a case one would be acting contrary to one's conscience. Moreover, this is what Kant ought to conclude; as pointed out in the introduction, it is difficult to see how Kant could avoid saying that if an agent acts contrary to his conscience, he has not done all that he ought as far as morality is concerned. The position that Himmler cannot act rightly no matter

what he does is the position Kant must take, and it is a position that (I think) coincides with intuition. This does not mean that Himmler would be praised if he had not helped the hypothetical Jews to escape; it means that he would have been blameworthy either way (even if for different things).

To put this another way, the Kantian insight here is that Himmler has put himself in a position in which he cannot act entirely autonomously. If he acts contrary to his conscience, then he will not be acting according to his principles, and if he does not act according to his principles, he will not be acting autonomously. But if Himmler's principles themselves are not universalizable, then even if he does act according to his principles, he will not be acting autonomously. It might look like there are two different standards being employed here, but really there are not. Whether a given principle is universalizable is dependent to some extent on a given agent's beliefs about how the world works and on whether the principle is consistent with his prior principles. There seems to be very little doubt that the Nazi ideology cannot be universalized or adopted autonomously. I shall return to this idea about the Nazi ideology below. If this is the case, Himmler's dilemma should be clear. A maxim to act against one's principles cannot be universalized (no matter what the principles);[20] a maxim to act according to Nazi principles cannot be universalized; and performing a good action despite having atrocious principles does not change the fact that one has atrocious principles. Himmler is in a lose-lose situation; he will lose for different things, but he will lose no matter what he does.

For now, this is the second point. Insofar as inverse akrasia is taken to describe agents who fail in their resolutions to act

[20] This raises an interesting and important point about moral conversion. However, I cannot discuss that point here.

subjectively rightly and, hence, act subjectively wrongly but objectively rightly, Kant's theory requires neither supplementation nor modification. Kant would not condemn such agents for the actions they perform (in Himmler's case, helping a Jewish family to escape) as much as for the principles they have adopted (in Himmler's case, to eradicate the Jews). Someone might argue (*pace* Kant) that a conscience that tells one to do wrong is simply a bad thing that one ought to disobey. But this objection falls wide of the mark. For Kant, conscience does not tell one what to do; it judges whether one is blameworthy. Once one sees that, for Kant, it is reason by means of understanding and the faculty of judgment that tells one what to do, the strength of Kant's position becomes clear. It is not that there is no other game in town. It is that the other game is a losing one. To borrow a turn of phrase from Kant's theoretical philosophy, principles without action might be empty, but action without principles is blind.

II. Acting contrary to a fallible judgment

In the introduction to this paper, I pointed out that Kant thinks that agents can be mistaken in their judgments as to whether something is a duty. In the previous section, I examined the strength of Kant's position when an agent's mistake is located in his understanding: he has bad principles. However, the mistake could come from elsewhere: the mistake could be located in an agent's faculty of judgment. Kant does think that the faculty of judgment is fallible, and he thereby seems to allow room for the possibility that agents make mistakes about whether actions are subjectively right, in accordance with the principles to which they (really) are committed. If this is so (if Kant really does allow room for this and if, in so doing, he is correct—if agents *can* make mistakes about whether a

given action is subjectively right) then this will be reflected in the final verdict reached by conscience. For example, if an agent mistakenly judges X to be subjectively right and acts accordingly, he will not be judged blameworthy by conscience even if he is acting both subjectively and objectively wrongly.[21] In other words, an agent might act according to conscience but (nonetheless) perform an action that is subjectively and objectively wrong.[22]

Kant's theory of conscience seems to force him to say that even if an agent is acting subjectively wrongly and objectively wrongly, if he is acting according to his (fallible) best judgment and, thus, according to his conscience, then he has done all that he ought as far as morality is concerned. This might seem counterintuitive at first. If an agent is a bad deliberator, then there seems to be very little reason why he should follow his best judgment. Moreover, if he acts both subjectively and objectively wrongly in following his best judgment, it is not clear how to make intuitive the idea that he has done all that he ought as far as morality is concerned. In order to illustrate this, consider a person with a car, A, whose principles are good as far as following the speed limit goes. That is, A knows that he ought never to drive above the speed limit and he fully intends to abide by this principle. Perhaps A even has prudential

[21] Alternatively, an agent might judge X erroneously to be subjectively right even though X is subjectively wrong and objectively right; he might judge X erroneously to be subjectively wrong even though X is subjectively right and objectively right; and he might judge X erroneously to be subjectively wrong even though X is subjectively right and objectively wrong. In this paper I am concerned mainly with the possibility described in the text above: acting both subjectively and objectively wrongly.

[22] Cases like this are used by Arpaly and Schroeder to motivate what they call a "whole self " theory ("Praise, Blame and the Whole Self"). Arpaly's and Schroeder's basic idea is that whether an agent is praiseworthy or blameworthy depends on whether the action that he performs is subjectively right (which is determined by whether it is in accordance with his "whole self ") rather than merely in accordance with his (fallible) best judgment (in accordance with his "reason").

reasons for never driving above the speed limit. Perhaps A has a hybrid and is a "hyper-miler," someone who takes the endeavor of trying to maximize gas mileage to the extreme. A accelerates so slowly (to ensure that his gas mileage remains above par) that it would be well nigh impossible for him to break the speed limit under normal conditions.

Now A is out driving late one night. During his midnight snack, he realized that there was no more milk, so he decided to make a quick trip out to the grocery store before bed. But he also realized that the grocery store he usually patronizes is not open so late at night. So he took care to look up the nearest Safeway (open 24 hours) and to write down directions before setting off. On the way there, A is on a long stretch of road with a 25 miles-per-hour speed limit. A has been on the road long enough to get his car up to 35 miles per hour and he is cruising along, happily observing that he is currently getting more than 50 miles to the gallon. Just then he sees the speed limit sign; there is a 25-miles-per-hour speed limit. Being a hyper-miler, A is neurotic about watching his speedometer; it is, after all, immediately adjacent to the real-time read-out showing his gas mileage. However, A is not accustomed to thinking about the speed limit; given his car and given his driving habits, one can see why not. Not being accustomed to this kind of thinking, it is perhaps unsurprising that A makes a mistake. His thought process is roughly as follows: "the speed limit is 25 miles per hour, I am going 35 miles per hour therefore I am driving below the speed limit."

If somebody had challenged A at that moment, telling A that he is speeding, A would have looked incredulous. He even might have fired back, "what do you mean? The speed limit is 25 miles per hour; I am only going 35 miles per hour."

It is not that A is not paying attention or that A is negligent or anything of that nature. On Kant's theory, one cannot infer negligence from the fact of error and one ought not to do so here. It is tempting to say that this sort of thing never happens or that it is not reflective of an agent's best judgment.

But it really does happen. Some agents are simply bad deliberators (the "wiring" is "awry"); some agents are good deliberators but have temporary "lapses." Moreover, having argued already that Kant does not preclude the possibility that an agent's best judgment is incorrect (and, thus, that an agent might act in accordance with his best judgment despite acting both subjectively and objectively wrongly), there does not seem to be any way of closing the door on cases such as A's.

I think that there are three things that Kant would have to say about such cases in general. The first has to do with the distinction between legal duties (duties of right) and moral duties (duties of virtue). The second has to do with the distinction between forming a judgment (in this case, the judgment that it is permissible to drive at 35 miles per hour) and acting accordingly (in this case, actually driving at 35 miles per hour). The third has to do with the judgment itself.

Assume for the moment that Kant must say that A has done all that he ought as far as morality is concerned (at least with regard to A's driving on the stretch of road with a 25-miles-per-hour speed limit). That is, assume for the moment that Kant's theory of conscience really does force him to conclude that A is morally blameless with regard to his driving. Given this assumption, it does not follow that Kant is forced to say that A cannot be held legally accountable for his speeding or that A ought not to be punished for breaking the law. Let me explain.

It is useful here to introduce the concept of strict liability. Strict liability refers to legal responsibility for which *mens rea* ("guilty mind") does not have to be proven in relation to one or more elements comprising the *actus reus* ("guilty action"). Strict liability laws were formalized in the 19th century to improve working conditions in factories, for it was found to be very difficult to prove *mens rea* in existing circumstances. The only defense available in a case of strict liability is due diligence; in such cases, the defendant must show beyond a reasonable doubt that he took every reasonable precaution (the normal standard of care is not sufficient).[23] Strict liability is found in civil law (for example, product liability and care of animals) and also in criminal law (for example, certain statutory offenses). Strict liability laws vary from legal code to legal code. However, a few general examples will help to illustrate the concept. If B sells alcohol to C in the USA, B can be found liable regardless of whether B believed that C was old enough to buy alcohol. Indeed, B can be found liable even if C showed B a fake ID that (1) misrepresented C's age and that (2) B reasonably believed to be genuine. The court must show merely that the liquor was sold to a person who was not, in fact, old enough to buy alcohol. Similarly, someone can get a speeding ticket in the USA even if he reasonably believed that he was driving within the speed limit. Finally, in most jurisdictions in the USA, keepers of animals are strictly liable for damages resulting from the trespass of those animals on someone else's property.[24]

[23] Some countries, such as Canada and India, have an additional category called absolute liability. Absolute liability is sometimes confused with strict liability. However, in countries that have both categories, the distinction is clear. Absolute liability does not allow any defense (not even due diligence).

[24] Sometimes the law contains exception clauses for dogs and cats. But other domesticated animals, such as cows and sheep, do not seem to enjoy this privilege.

The point of bringing in the concept of strict liability is that it illustrates that it is possible to ascribe legal responsibility in the absence of ethical culpability. That is, regardless of the agent's state of mind, his principles or anything else, he can be held strictly liable for his actions. It would be too difficult to enter into a discussion of Kant's doctrine of right in this paper. However, it is worth pointing out that some prominent commentators believe that, on Kant's account, all legal duties are strictly liable.[25] Thus, it is not only logically possible for Kant to claim that agents are legally culpable for their actions even if they are morally blameless, it is, according to some commentators, very probable that he really would do so. This is the first point. Kant can say (and, according to some commentators, really would say) that A, moral blamelessness notwithstanding, is legally culpable and deserving of punishment for speeding.

This is an important point. In thinking about A's case (and others like it), one must disambiguate one's intuitions about whether A is legally culpable and ought to be punished from one's intuitions about whether A is morally culpable. One might feel sorry for A but think that A needs to be more careful, alerted to the fact that he is not a flawless deliberator. Alternatively, one might think that A is morally inculpable in this instance but worry that there does not seem to be any good way to distinguish a case in which A really does make a faulty judgment despite having good principles from a case in which A makes a good judgment but has bad principles. One might admit that people sometimes make morally inculpable mistakes about whether they are driving under the speed limit, but one might think that punishing such mistakes is a

[25] See, for example, Arthur Ripstein, *Force and Freedom: Kant's Legal and Political Philosophy* (Cambridge, Harvard University Press, 2009), and personal communication.

necessary evil, required in order to keep people who regularly make such mistakes off the road. An agent might be bad at calculating and there might be nothing morally inculpable about this. All the same, such an agent should be kept away from activities in which his poor calculating abilities might result in the death of innocents. Finally, one might think that Kant's views of legal culpability are totally bogus; one might find the notion of strict liability pernicious. But that is another issue. The point is merely that in thinking about such cases, one must distinguish intuitions about legality from intuitions about morality.

The second point, which has to do with the distinction between the formation of a judgment and acting on it, gets more into the meat of the issue. This distinction might be easier to see in other cases. For example, suppose that D and E have gone out to eat together. As per usual, D pays for the meal and E picks up the tip. E plans (as usual) to give a tip of 23 percent; E has a thing for prime numbers, but 19 is too little and 29 is too much. In the process of doing the multiplication, E forgets to carry a 1 and the tip is quite a bit lower than it ought to be. Out of habit, E passes the receipt to D to check it; D has no calculator and follows in E's tracks, not catching the mistake. D gives the thumbs up, passes the receipt back to E and prepares to go. E rounds up to the nearest dollar amount, sets the money on the table and leaves with D. I think that consideration of such cases reveals that Kant's theory gets things exactly right.

One might not approve of A's speeding or of E's small tip. But in these cases, it is not the actual speeding or the giving of the tip that is a problem; it is the judgment itself. Insofar as the agent acts in accordance with his best judgment and, thus, in accordance with conscience, it really is intuitive to say that the agent has done all that he ought as far as morality is concerned.

But this does not mean that he is not to be faulted for the way in which he actually formed the judgment. An agent who calculates the product of 100 and 0.2 to be 10 rather than 20 and acts accordingly is to be faulted (if at all) for his miscalculation rather than for his conduct. Of course it would be better if he had calculated 20 and acted accordingly, but if he calculates 10 and acts as if he calculated 20, then he is compounding rather than correcting the error.

One might worry that a Kantian theory cannot sustain this conclusion. Above I defended the idea that an agent who acts against his principles has not done all that he ought as far as morality is concerned by appeal to the notion of heteronomy. That is, I argued that such an agent would be acting against his own principles and, hence, heteronomously and, thus, immorally as far as Kant is concerned. But if an agent's best judgment is incorrect and he judges incorrectly that X is in accordance with his principles, then it looks like this defense is undercut; such an agent, it seems, really ought to go against his best judgment, for only by going against his (in this case faulty) best judgment will he be able to act autonomously. I argued above that Kant's theory of conscience might need an exception clause in cases in which agents have diseased consciences. But perhaps it requires an exception clause for any time that an agent's best judgment is faulty. If so, then Kant's claim about acting in accordance with conscience has nothing to do with the judgments of conscience as such.

What this worry overlooks is that there is something special about one's best judgment. That is, it makes no sense to advise an agent to act against his best judgment. An agent's best judgment is like the judgment in a court of law. One can appeal the judgment and one can disagree with the judgment, but unless the court system is corrupt (the agent's conscience is diseased), one cannot avoid the

fact that the judgment is legally binding; that is what it is for it to be the judgment in a court of law. Just so, it never can be rational or autonomous to act against one's best judgment. If one's judgment is flawed, it might turn out to be the case that one is in a lose-lose situation; acting against one's best judgment is bad and so is acting in accordance with it. But there is no winning game in acting against one's best judgment. That, anyway, is what I take Kant's position to be, and it is sustainable even if some might find it unpalatable.

This point should not be misunderstood. Suppose that according to F's best judgment, he ought to break his promise to G. It makes sense to say that F's best judgment is wrong in this case, to say that F ought not to break his promise to G. It makes sense to argue with F about this, to try to convince him that really he ought to keep his promise to G. In doing so, one would be trying to convince F that his best judgment is mistaken and that he ought to reconsider, to come to a new best judgment. The point is that it does not make sense to tell F that he ought to act against his best judgment *simpliciter*. This is because one cannot come to the conclusion (as a result of deliberation) that one ought to act against the conclusion of one's deliberation; an agent whose best judgment is, "I ought to act against my best judgment," is not making sense; this judgment is as absurd as the command, "never do anything I command." Moreover, adding an exception clause ("I ought to act against my best judgment except for this one") simply makes the judgment into an admission that one's cognitive faculties are so faulty that one ought not to be considered a rational being.

It makes sense to say that an agent's best judgment is wrong in any given case and to say that it would have been better if the agent had not acted on his best judgment. But these are different things from advising the agent to act against his best judgment. Moreover, the fact that acts of conscience are essentially self-

reflective makes the question of what makes sense for an agent to be advised to do more relevant to issues about conscience than the question of what we think the agent should have done or what we think would have been best, all things considered, for the agent to have done. To advise an agent to act against his best judgment is like advising an attorney to appeal a decision of a court of last instance; it is contradictory to suppose there could be such an appeal. That is the point of saying that it is a court of last instance.

The third and final point is an extension of the second point. One might argue that some judgments are so basic or so important that one cannot be in error about them inculpably.[26] It should be open to a Kantian to say that some errors of judgment are culpable errors because they are errors not merely about what one's duty is but errors necessarily affecting the inner judicial process of one's conscience. In other words, if one never submitted oneself to a genuine examination of one's own culpability but rested content in a dogmatic fashion with one's grossly erroneous view of one's duty, then there is no real possibility for acting in accordance with one's conscience and, therefore, one's best judgment. It is unlikely that one will say this about A. Driving 35 miles per hour in a 25-miles-per-hour zone is probably not considered by most to be such a horrible thing. But depending on how one fills out the description of A, one might think that A never even formed the judgment that it is permissible to drive at 35 miles per hour; one might think A was operating on automatic pilot and that the speed limit never really registered fully in his conscious awareness. But if so, then A never employed his conscience and, hence, it is not the case that A was

[26] Kant himself discusses this possibility in his *Lectures on Ethics*. For a discussion of this, see Samuel Kahn, "The Interconnection of Willing and Believing in Kant's and Kantian Ethics," *International Philosophical Quarterly* (forthcoming).

acting in accordance with his conscience and, hence, the problem vanishes.

It might be tempting to say that errors in basic arithmetic are of this nature. That is, it might be tempting to say that judging that X is less than Y or that the product of X and Y is Z are examples of judgments so basic that no agent can be in error inculpably about them; if someone makes an error about such things, negligence can be inferred. I want to resist this temptation. I do not want to resist this temptation merely because it would render the examples I gave of A and E cases of negligence rather than cases of inculpable false judgment. I want to resist this temptation because I think that agents regularly make errors in basic arithmetic and I think that much of the time such errors are inculpable.

However, I think that there are some contexts in which one cannot judge inculpably that some one adult human being or group of adult human beings are not rational. Of course, not all adult human beings are rational; surely adults with full-blown dementia are not rational beings. Moreover, there are borderline cases; legally, citizens in the USA reach majority at 18, but surely there is some hazy gray zone in individual cases. Finally, one must make allowances for culture and background; perhaps some judgments inherent in the institution of slavery in ancient Greece were OK even though they were certainly not in the antebellum South. But the point remains: there are some contexts in which one cannot judge inculpably that some one adult human being or group of adult human beings are not rational. One example of this might be the judgment in Nazi Germany that all Jews are (by virtue of being Jewish) not rational beings. That is, one might think that there is no way that someone inculpably could come to believe the Nazi ideology. Perhaps that is even part of the point of calling it an ideology.

To sum up: In this section I considered the possibility of an agent who acts in accordance with his fallible best judgment and, thus, subjectively and objectively wrongly. I argued that one must distinguish between one's intuitions about legal culpability and one's intuitions about moral culpability; if on Kant's account agents are strictly liable for failing in duties of right, then acting in accordance with one's best judgment would not be any more exculpatory than acting in accordance with one's principles as far as legality is concerned. One can be culpable for failing to fulfill a legal duty (and, hence, punishable) regardless of whether one acted in accordance with one's best judgment. I argued that Kant might be forced to say that an agent is morally blameless for the action that he performs in accordance with his best judgment (even if the resulting action is neither subjectively nor objectively right), but this does not preclude saying that the agent is blameworthy for the judgment itself. Depending on how the agent arrives at the judgment in question, he might be judged blameworthy. For example, if he arrives at his "best" judgment negligently or without due reflection, he might be judged blameworthy for that fact. Third, there are some judgments that are so basic or so important that Kant could say that an agent cannot be inculpable if he makes them erroneously. This could be motivated by saying that there are some judgments so basic that no error about them could count as a judgment (or no agent who makes such an error could count as an agent).