



BRILL

Journal of Moral Philosophy 5 (2008) 327–360

JOURNAL OF
MORAL
PHILOSOPHY



www.brill.nl/jmp

Unstable Autonomy: Conscience and Judgment in Kant's Moral Philosophy*

Dean Moyar

Department of Philosophy, Johns Hopkins University
3400 N. Charles St., Baltimore, MD 21218, USA
dmoyar@jhu.edu

Abstract

In this paper I argue that Kant's claims about conscience in his moral writings of the 1790s reveal a fundamental instability in his moral philosophy. The central issue is the relationship between the moral law as the form of universality and the judgment of individuals about specific cases. Against Thomas Hill's claim that Kant has only a limited role for conscience, I argue that conscience has a comprehensive role in Kantian deliberation. I unpack the claims about conscience in the *Metaphysics of Morals* to show that they describe conscience as both a basic act of self-consciousness and as an all-things-considered judgment. I outline the role of conscience in moral motivation, and argue that taken together Kant's writings about conscience reveal a way to rethink Kant's conception of the Fact of Reason.

Keywords

Kant, Moral Philosophy, Conscience

In his 1798 'System of Ethics', J.G. Fichte writes that the Universal Law formula of Kant's Categorical Imperative is only *heuristic*, not *constitutive* of morality. His main reason for this claim has to do with the nature of moral judgment. Fichte gives Kant's formula as 'Act in such a way that you could think of the maxim of your will as a principle of a universal legislation', and writes of it,

It is by no means a principle [*Prinzip*], but only a consequence of or an inference from a true principle, that is, a consequence of the command concerning the

* Many people have given helpful comments on this paper. I would especially like to thank Eckart Förster, Sean Greenberg, Mark Jenkins, Joseph Schear, Oliver Thorndike, and Allen Wood for their criticisms and suggestions.

absolute self-sufficiency of reason. The relationship in question is not that something ought to be a maxim of my will because it is a principle of a universal legislation, but rather the converse – because something is supposed to be a maxim of my will it can therefore also be a principle of a universal legislation. The act of judging comes purely and simply from me [*geht schlechthin von mir aus*]. This point is also clear from Kant's proposition, for who is it that judges in turn whether something could be a principle of a universal legislation? This is surely I myself.¹

This is a classic example of Fichte taking up the 'spirit' of Kant's philosophy while pushing against a natural way to read the 'letter' of Kant's writings. In this paper I address the instability in Kant's view of moral judgment that Fichte draws out in this passage. The instability arises because, as Fichte puts it here, two aspects of Kant's conception of moral freedom can be separated: 'the absolute self-sufficiency of reason' and the 'principle of universal legislation'. Fichte is claiming that practical reason and the pure act of the I that is the root of reason's self-sufficiency are not constituted by the universality of the Categorical Imperative. In maintaining that 'The act of judging begins purely and simply with me', he is in effect giving priority to the self over the universal form of lawfulness, and bringing out into the open the latent instability in Kantian autonomy.

Kant's writings on conscience are the primary conceptual site of the tension between the law and the self in his moral philosophy. From the first extended discussion of conscience in the 1791 'Theodicy' essay, to the two treatments of conscience in the 1797 *Metaphysics of Morals*, Kant's treatment of conscience shows a remarkable lack of uniformity.² The continually shifting descriptions of conscience in Kant's late moral philosophy reflect a tension in his view of autonomy between the authority of the moral law as a function of the understanding (i.e. the faculty of rules) and the moral law as a function of self-imputation, of giving the law *to myself*. Fichte himself holds that conscience is an 'absolute criterion' for the rightness of our moral judgment, and he claims that we can infer from our awareness of the moral law an original self-constituting act. My aim in this paper is to show that Kant's reflections on conscience led him towards a Fichtean view of the primacy of a pure act of

¹ J.G. Fichte, *Johann Gottlieb Fichtes sämtliche Werke*, ed. I.H. Fichte (Berlin: Walter de Gruyter, 1971), vol. IV, p. 234; *The System of Ethics*, ed. and trans. Daniel Breazeale and Gunther Zöller (Cambridge: Cambridge University Press, 2005), pp. 222–23.

² One task that I am not taking on in the present paper is a thorough review of the development of Kant's thinking on conscience. For such a review, see Gerhard Lehmann, *Kants Tugenden* (Berlin: Walter de Gruyter, 1980), and Thomas Hoffman, 'Gewissen als Praktische Apperzeption: Zur Lehre von Gewissen in Kants Ethik-Vorlesungen', *Kant-Studien* 93 (2002), pp. 424–43.

self-consciousness and the irreducibility of the individual's authority in moral judgment to the universal law.

Many Kantians will see no tension or instability here at all, for a very basic move in Kant's moral philosophy is to identify the moral self with the form of the universal law. Kant was very much concerned to separate questions about freedom and the a priori justification of the moral law's authority from questions about the application of the moral law in particular actions and judgments. This move allows Kant, in his foundational writings of the 1780s, to avoid difficult issues in adjudicating the authority of the universal law and individual judgments in cases of moral conflict. Kant avoids coming to terms with the tension between the law and the self by treating moral action in the abstract and taking relatively simple cases as conceptually central. Attending to the exigencies of moral judgment, on the other hand, means coming to terms with an authority for the self that is not reducible to the universal law. Though Kant is in one sense very friendly to the first-person perspective, his conception of moral justification rests entirely on the form of lawfulness.³ I will show that Kant's claims in the 1790s about conscience as a capacity for the self-imputation of actions put pressure on the identification of the law and the self in judgment. Detaching my authority in judgment from the universal form of lawfulness does not mean abandoning autonomy as the basis for moral philosophy. As expressed in the quote from Fichte above, the irreducibility of the authority of judgment to the principle of lawfulness need not displace *reason* from its basic authority. But it does imply that the moral law of reason is something more basic than the universal form of lawfulness.

Before turning to Kant's text, a few initial comments are in order about the concept of conscience. Assuming that conscience is not a mysterious oracular source of moral truth (as on some traditional theistic views), and that conscience is not a mere sponge of whatever norms are accepted in one's social environment, a theory of conscience should be seen as a theory about the nature of the claim to authority for an individual's judgment. The distinctive mark of this claim is that it is made *in the first person*. As such, it shares the essential tension of the first-person standpoint, and of the indexical 'I',

³ Christine Korsgaard defends Kant's equation of selfhood and lawfulness as a thoroughly first-personal account in Christine Korsgaard et al., *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996). Her commentators all in one way or another take issue with her first-personal reading of Kant, trying to drive a wedge between individual self-consciousness (or 'identity') and the lawfulness of reason. I think Korsgaard's aim is the right one, though I think that her commentators are right to say that her view sits uneasily with Kant's commitment to strict lawfulness. The view that she wants to find in Kant could be drawn out better by attending to what he says about imputation in terms of conscience.

a tension between a universal use (this ‘I’ could be anyone’s) and a particular use (I am talking about *me*). This tension is closely related to the central issue of this paper. On the one hand, the authority of conscience might be reducible to the universal law, so that in referring to the authority of conscience I am simply claiming to judge according to the law. On the other hand, it seems that I only need to invoke conscience when the authority of universal principles has run out, so to speak, so that the particularities of my character, and my sense of my own integrity or identity, come to be decisive. A central aim of my analysis of Kant’s writings on conscience is to show that the reference to the authority of the self in difficult moral judgments cannot be viewed just as a question of application. In order to secure the *rationality* of such judgments we need to rethink along Fichtean lines the very foundation of moral obligation.

1. Conscience and the Fact of Reason

Kant’s writings on conscience have not played a major role in the revival of interest in his moral theory.⁴ One reason for this neglect is that Kant himself only assigns conscience an official place in his theory after his two most well-known works in the ethics, the *Groundwork* and the *Critique of Practical Reason*. Yet one can read the most important shift from the *Groundwork* to the *Critique of Practical Reason*, Kant’s abandonment of the attempt to deduce the moral law and his introduction of the notorious ‘Fact of Reason’ (abbreviated as *Factum* in the following) argument, as a shift that brings conscience to the fore. The *Factum* is a claim about the unavoidability of recognizing the authority of the moral law, and thus about the undeniability of freedom from the practical point of view. The *Factum* is a claim that is naturally read as a claim about the conscience of any moral agent – my conscience forces me to recognize the authority of the law over my judgments, and it tells me I can take the

⁴ Thomas Hill has been the main exception. Allen Wood also has included a lengthy discussion of conscience in Allen W. Wood, *Kantian Ethics* (New York: Cambridge University Press, 2008). Korsgaard and Barbara Herman both mention conscience within their discussions of more familiar aspects of Kant’s theory, but neither of them explores Kant’s own comments on conscience. On the theme of conscience and Kant one could also consult J. David Velleman’s ‘The Voice of Conscience’, *Proceedings of the Aristotelian Society* 99 (1998), pp. 57–76. In Velleman’s case, however, the focus is not on what Kant wrote about conscience, but rather on how we can think of the more familiar Kantian apparatus (duty, moral law) in terms of conscience. My claim, by contrast, is that if we take Kant’s specific statements about conscience seriously we will gain insight both into his moral theory and into our ordinary views of conscience.

moral law as the determining ground of my will over and above all non-moral motives. Such an identification of the *Factum* and conscience would be too quick, but it does provide a useful frame and starting point for Kant's attempts to come to terms with how the moral law operates in judgment and willing. In this section I consider the basic nature and function of the *Factum* and how it could be read as a claim about the authority of conscience. After examining Kant's writings about conscience in the 1790s in the central sections of the paper, I will conclude by claiming that Kant arrived at a concept of conscience that could do most of the work of the original *Factum* while avoiding the main pitfalls of the original account.

The strategy of Kant's deduction from the *Factum* is to infer from this subjective consciousness to the reality of freedom as the actual ability to determine the will independently of all inclination.⁵ The *Factum* must be read as a certain kind of consciousness, and is best expressed in this passage:

Consciousness of this fundamental law may be called a fact of reason because one cannot reason it out from the antecedent data of reason, for example, from consciousness of freedom (since this is not antecedently given to us) and because it instead forces itself upon us of itself as a synthetic a priori proposition... However, in order to avoid misinterpretation in regarding this law as *given*, it must be noted carefully that it is not an empirical fact but the sole fact of pure reason which, by it, announces itself as originally lawgiving (*sic volo, sic jubeo*).⁶

It is *through* the moral law that we first become conscious of freedom, so the attempt in *Groundwork* III to derive the moral law from freedom fails. This is why Kant writes of the moral law as the *ratio cognoscendi* of freedom.⁷ The *Factum* is not an intuition of a *given* empirical fact, but rather consciousness of the ability of the will to determine itself a priori through the moral law. The goal of the deduction of freedom in the *Critique of Practical Reason* is to understand how the *Factum* can underwrite the inference to transcendental freedom as the a priori determination of the will by law-giving reason.

If one thinks of conscience as the awareness that one ought to act according to the moral law, it is easy to see how one could think that the *Factum* is just

⁵ There are passages in which Kant identifies the *Factum* with the moral law itself. But as Henry Allison has stressed, the point of the 'Analytic' of the second *Critique* is precisely to establish the objective reality of the moral law (and freedom), so it would simply beg the question if the *Factum* were taken as the moral law. Henry Allison, *Kant's Theory of Freedom* (Cambridge: Cambridge University Press, 1990), pp. 231-33.

⁶ I cite Kant's texts from the Academy edition: *Kant.Ak.* = *Gesammelte Schriften*. Königlich Preussischen Akademie der Wissenschaften (Berlin: 1900–). *Kant.Ak.* 5:31; *Practical Philosophy*, ed. and trans. Mary J. Gregor (Cambridge: Cambridge University Press, 1996), pp. 164-65.

⁷ *Kant.Ak.* 5:4n; *Practical Philosophy*, p. 140n.

a description of what we ordinarily refer to as conscience. If everyone can identify the *Factum* with the operation of conscience, that could even serve as evidence for Kant's claim that freedom can be inferred from common moral consciousness. The *Factum*, as the consciousness of the moral law as authoritative for the will of a rational being, has two practical components, namely judgment and motivation. The *Factum* is the agent's awareness that she is able to actually judge according to the principle of the moral law, and that the law must be effective in producing the right kind of motivation. If the agent is conscious in moral reasoning and action of the effectiveness of the moral law in the *judgmental* and *motivational* dimensions, then a case can be made for the reality of that law and of freedom. Kant's deduction of freedom thus naturally includes analyses of the moral law as a principle of judgment and of the 'incentive' of morality.⁸ In the section 'Of the Typic of Pure Practical Judgment', Kant gives a version of the universal law formulation of the Categorical Imperative that is a 'rule of judgment'. This formulation draws on the concept of a 'law of nature', 'ask yourself whether, if the action you propose were to take place by a law of the nature of which you were yourself a part, you could indeed regard it as possible through your will'.⁹ In the third chapter of the 'Analytic', Kant gives the motivational component with his doctrine of respect, a 'moral feeling' that proceeds from the representation of the law. The feeling of respect checks self-love and 'strikes down' self-conceit so as to clear a motivational space for the law itself to serve as the incentive of my will. The affirmation of both these elements in moral reasoning and action is the *Factum* that is the key premise in the deduction of freedom.

In a recent article Ian Proops has claimed that Kant tried to defend the *Factum* argument on the grounds that it accurately describes the conscience that everyone in fact possesses. While such a direct connection between the *Factum* and conscience is doubtful as a description of what Kant thought he was doing, it does provide a point of entry into considering how conscience could play a foundational role in Kant's ethics. Proops argues that the *Factum* is designed to show 'the *purity of origin*' of the moral law, and, drawing on the nature of legal deductions (building on work by Dieter Henrich), that this *Factum* admits of a proof.¹⁰ As evidence that Kant has such a proof in mind,

⁸ The deduction thus encompasses the whole of the 'Analytic'. On this point I am in agreement with Paul Franks, *All or Nothing: Systematicity, Transcendental Arguments, and Skepticism in German Idealism* (Cambridge, MA: Harvard University Press, 2005), chapter 5.

⁹ *Kant.Ak.* 5:69; *Practical Philosophy*, p. 196.

¹⁰ The discussion comes at the end of an article that deals primarily with the question of 'fact' in the deductive strategy of the first *Critique*. Ian Proops, 'Kant's Legal Metaphor and the Nature of a Deduction', *Journal of the History of Philosophy* 41.2 (2003), pp. 209–29, p. 225.

Proops cites the passages in which Kant refers to the undeniability of the *Factum* as stemming from what *all agents* actually do in judging morally and appraising their actions.¹¹ Proops raises an objection to this ‘proof’ in terms of conscience, and cites Kant’s 1784 *Lectures on Ethics*, where Kant considers the objection that ‘conscience is a product of art and education’.¹² Proops reads Kant’s insistence that one cannot escape the pangs of conscience as supporting the proof of the *Factum*.¹³ This constancy is not just a matter of using the moral law as a rule of judgment, but is also a constancy in the responsiveness to the law through a feeling of respect that cannot be attributed merely to one’s social environment. Proops’s interpretation is instructive because he takes Kant’s *Factum* argument to depend on evidence of unanimity of conscience across various educational backgrounds. Although Kant does appeal to unanimity in the passages that Proops cites, such a proof strategy is not very promising for bolstering the *Factum* argument. The obvious problem with the *Factum* argument is that it seems to replace a deduction, however problematic, with a simple assertion. Though Proops improves on the *Factum*’s status by showing that Kant thought a proof of the *Factum* was available (namely the unanimity of conscience), this ‘proof’ either amounts to an assertion of a fact of human nature, or something to be established by empirical study. The latter option, as so much psychology and anthropology of the last two centuries has shown, is simply untenable.

In another recent account of the *Factum* argument, Paul Franks foregrounds the skeptical objections to Kant’s argument, including skepticism about the argument that the deduction of freedom can rely on the actual unanimity of consciousness of the moral law. Franks instead reads the *Factum* as a first-personal *act*, an avowedly Fichtean reading that provides an alternate connection of the *Factum* and conscience. He stresses the motivational dimension of the *Factum* through highlighting Kant’s example of the choice of the gallows or an immoral action. Franks calls the *performative* dimension of the deduction the reliance on what happens when we read the text: in considering the

¹¹ Thus, just after the passage I cited above, Kant writes, ‘The fact mentioned above is undeniable. One need only analyze the judgment that people pass on the lawfulness of their actions ...’ *Kant.Ak.* 5:32; *Practical Philosophy*, p. 165.

¹² *Kant.Ak.* 27:355–56; *Lectures on Ethics*, ed. Peter Heath and J.B. Schneewind, trans. Peter Heath (Cambridge: Cambridge University Press, 1997), p. 134.

¹³ ‘What would reveal the non-empirical origin of the idea of duty would be its constancy across persons whose quality and level of moral education differ widely’. Proops, ‘Kant’s Legal Metaphor’, p. 226. He also writes, ‘What I have identified as the proof of the *factum* is not a way of grounding or justifying the moral law. It is not a way of demonstrating that we ought to follow it, but rather an argument that what we take to be our obligations – the deliverances of conscience – are indeed genuine manifestations of the moral law’. Proops, ‘Kant’s Legal Metaphor’, p. 228.

example ‘we actually *produce* the feeling of respect. So Kant is claiming that in reading the Analytic, we *demonstrate* the reality of freedom by *producing* an effect necessitated by the moral law’.¹⁴ In doing so ‘we *actualize* our practical freedom and so demonstrate that we are really free’, which forecloses the possibility (for me) that my moral consciousness might just be an illusion.¹⁵ Thus the deduction has a ‘peculiarly *first person singular* character’, so that ‘Moral belief has universal validity, but because it rests upon subjective grounds...it is unlike theoretical belief and can never become *knowledge*’.¹⁶ Franks claims that the deduction of freedom is only valid *for the reader* who successfully takes up and is moved by the examples considered in the text. Franks thus avoids the problems that arise with having to claim that everyone does as a matter of fact have the same affectively laden consciousness of the moral law. But the cost of such a first-personal reading is unacceptably high. Kant surely meant for the deduction to establish the reality of freedom from *the* practical point of view, not just from *my* practical point of view. Franks thinks that Kant does not or cannot simply claim that *any* moral agent would in fact thus respond to the example, but Kant needs a version of this claim. He needs to be able to say that this responsiveness is constitutive of moral agency, is necessary to be a moral agent at all. Such a necessary responsiveness is just what Kant’s theorizing about conscience in the 1790s aims to demonstrate.

Franks has taken an important step towards a more plausible *Factum* argument in reading it as an act. I interpret Kant’s theory of conscience in the 1790s as a theory of such an act of self-consciousness, a necessary first-person presupposition of moral judgment. In this act I take up the practical point of view, acknowledge my freedom and responsibility for the morality of my actions. The first-person characterization of the *Factum* brings it close to the act of self-imputation that Kant already identifies with conscience in the ‘Critical Elucidation’ at the end of the Analytic of the *Critique of Practical Reason*. He writes,

So considered, a rational being can now rightly say of every unlawful action he performed that he could have omitted it even though as appearance it is sufficiently determined in the past and, so far, is inevitably necessary; for this action, with all the past which determines it, belongs to a single phenomenon of his character, which he gives to himself and in accordance with which he imputes to himself, as a cause independent of all sensibility, the causality of those appearances.

The judicial sentences of that wonderful capacity in us which we call conscience are in perfect agreement with this.¹⁷

¹⁴ Franks, *All or Nothing*, p. 287.

¹⁵ *Ibid.*

¹⁶ Franks, *All or Nothing*, p. 294.

¹⁷ *Kant.Ak.* 5:98; *Practical Philosophy*, pp. 218–19.

The main point of this passage is that conscience confirms that I must consider myself as a noumenal agent. This is to consider myself from the standpoint of responsibility, or justification, rather than from the standpoint of explanation. As Kant goes on to say, conscience has an affective dimension as well, which in this passage he identifies as a painful feeling of remorse that does not subside with time. The agent ‘imputes to himself’ the causality of his actions, judging that he is responsible and connecting ‘feeling with it morally’.¹⁸

We can see from this mention of conscience that Kant links it, if rather obliquely, to the two elements of the *Factum*, namely motivation and judgment. We can also see in a preliminary way how conscience might improve on the foundational ideas in the *Factum* argument proper. In that argument, and in his moral philosophy generally, Kant treats *judgment* as subsuming a case under a moral principle or testing one’s maxim by the Categorical Imperative test to see whether it is contradictory. But as soon as multiple moral considerations are in play, it becomes hard to see the judgment as *constituted* by the Categorical Imperative. Take a garden variety case of moral conflict. What am I doing when I deliberate about stopping to help someone whose car is stalled on a lonely road late on a wintry night? Usually my obligation to keep my family in the car safe will outweigh the altruistic obligation to help. A similar problem arises with motivation. When there is no clear conflict between self-love and the moral law, but between competing obligations, it is unclear how the merely negative feeling of respect for the law really determines in such cases the proper motivation on which to act. I want to do the right thing, for the right reasons, and though Kant’s theory might help me rule out my desire to get to my destination before the telecast of the big game begins, it does not help settle the question of competing grounds of obligation. Moral agents make complex judgments all the time, sometimes easily and sometimes after much deliberation. The question that brings conscience to the fore in the 1790s writings is whether it can help guide those judgments.

2. ‘Moral Judgment Passing Judgment on Itself’

With his discussion of conscience in *Religion Within the Limits of Reason Alone*, Kant attempts to give an explicit systematic role in practical reason to conscience. But rather than align it with his foundational justificatory concerns, he tries to grant it only a restricted authority in applying the moral law.¹⁹

¹⁸ *Kant.Ak.* 5:99; *Practical Philosophy*, p. 219.

¹⁹ The first significant systematic discussion of conscience in one of Kant’s published texts comes in the 1791 essay ‘On the Miscarriage of All Philosophical Trials in Theodicy’. *Kant.*

He writes in the first line of the section entitled ‘Concerning the Guiding Thread of Conscience in Matters of Faith’, that ‘conscience itself can serve as guiding thread in the most perplexing moral decisions’.²⁰ It is very unclear from the discussion, however, exactly how conscience is supposed to help with problematic cases. Most of Kant’s discussion is about the need to be certain about one’s actions, and does not really address the role of conscience in reaching that certainty through deliberation. He begins by asking how there can possibly be such a ‘*consciousness which is of itself [für sich selbst] a duty*’, since ‘consciousness of all our representations seems to be necessary only for logical purposes, hence only conditionally, whenever we want to clarify our representation’.²¹ The important point here is that conscience is to make available to the subject ‘all our representations’, and so is the duty to achieve explicit clarity about what one has done in moral deliberation. For logical purposes I can analyze my perceptual judgment, for instance, to gain clarity about the role of the understanding in synthesizing the spatiotemporal manifold. In contrast to the theoretical case, in practical deliberation there is a demand that the process of judgment be transparent to the subject who is about to act. As Kant writes with reference to conscience in the ‘Theodicy’ essay a few years earlier, even if I have to admit that the truth of what I say is open to doubt, the *truthfulness* or sincerity of my judgment must not be an open question for me.²²

Kant’s detailed description of conscience in the *Religion* is especially valuable because it provides a sketch of the anatomy of moral reasoning and displays his reservations about the authority of conscience. He writes,

It is a moral principle, requiring no proof, that we *ought to venture nothing where there is danger that it might be wrong (quod dubitas, ne feceris!* Pliny). So the

Ak 8:267-70; *Religion and Rational Theology*, ed. and trans. Allen W. Wood and George di Giovanni (Cambridge: Cambridge University Press, 1996), pp. 34-36. I refrain from analyzing that passage on its own here, for all of its main elements come up in the discussions with *Religion* and the *Metaphysics of Morals*.

²⁰ *Kant.Ak.* 6:185; *Religion and Rational Theology*, p. 202. The discussion of conscience comes near the end of the book, in the context of an examination of acting on matters of faith. One could seek to limit conscience to such religious questions, but it is clear that conscience is operative in action in general, and that acting on faith is simply the most pressing case of the need for conscience. That is, conscience is always in play in moral action, but it is most necessary in those cases in which the reasons for action go beyond those basic moral intuitions Kant thinks are obvious to everyone.

²¹ *Kant.Ak.* 6:185; *Religion and Rational Theology*, p. 202.

²² *Kant.Ak.* 8:268; *Religion and Rational Theology*, p. 34. There Kant relates this point to the idea that an erring conscience is an absurdity, a point curiously absent from the *Religion* discussion (I pick up that claim in section 5). I suspect that this absence has to do with the context, in which Kant is mainly concerned to argue against the Inquisition and the supposed claim to infallibility of the authority of the Catholic Church. Introducing the idea that conscience cannot err would have seemed like a claim for infallibility in matters of faith rather than an argument against it.

consciousness that an action *which I want to undertake* is right, is unconditional duty. Now it is understanding, not conscience, which judges whether an action is in general right or wrong. And it is not absolutely necessary to know, of all possible actions, whether they are right or wrong. With respect to the action that *I* want to undertake, however, I must not only judge, and be of the opinion, that it is not wrong; I must also be *certain* that it is not wrong. And this demand is a postulate of conscience to which is opposed *probabilism*, i.e., the principle that the mere opinion that an action may well be right is itself sufficient for undertaking it. – Conscience could also be defined as *the moral faculty of judgment, passing judgment upon itself*, except that this definition would be much in need of prior clarification of the concepts contained in it. Conscience does not pass judgment upon actions as cases that stand under the law, for this is what reason does so far as it is subjectively practical (whence the *casus conscientiae* and casuistry, as a kind of dialectic of conscience). Rather, here reason judges itself, whether it has actually undertaken, with all diligence, that examination of actions (whether they are right or wrong), and it calls upon the human being himself to witness *for* or *against* himself whether this has taken place or not.²³

While in the second *Critique* passage Kant wrote of conscience as functioning primarily in the retrospective mode, assessing what *I have done*, here he insists that this certainty must be achieved prior to performing any action.²⁴ It is ‘unconditional duty’ to achieve the consciousness that the action I desire to perform is not wrong. The demand here can only be to achieve subjective certainty that I have exhausted all my deliberative resources in assessing a case. The sentence that begins ‘Now it is understanding...’ warns against thinking that it is up to the authority of individual conscience to decide what is right or wrong. Kant insists that the understanding determines objectively the actual rightness or wrongness of an act. This role for the understanding is, I take it, the *constitutive* role for the Categorical Imperative that Fichte denied in the passage cited in my opening. Kant’s use of the impersonal formulation, ‘understanding... judges’, represents his attempt to shield the moral objectivity of rules from the exigencies of moral decision-making in specific cases. While Kant concedes that there is no duty to know of all possible cases whether they are right or wrong, even here he does not say that they *cannot be* known as right or wrong (it would presumably just be a *very long* moral handbook). The rightness or wrongness of any action is supposed to be determinable by the understanding and reason in so far as they are distinct from the specific activity of conscience.

The ineluctably individual character of conscience comes out in the next sentence, which is among the most intensely first-personal in all of Kant’s moral writings. With the repeated ‘I’, Kant emphasizes that I am responsible

²³ *Kant.Ak.* 6:185-86; *Religion and Rational Theology*, pp. 202-203.

²⁴ In this paper I focus mainly on the prospective mode of conscience, for it is in deliberation about what to do that the most pressing questions about the role of conscience come up.

for knowing with certainty whether *this* action, here and now, is or is not contrary to morality. Conscience's special role within practical reason is thus to act as the last gatekeeper for a practical judgment becoming an intention to act.²⁵ The most obvious contrast with the 'opinion' that Kant aligns with 'probabilism' would be *knowledge* of the case. Would it make sense for conscience to secure justified true belief? That depends on what we think securing such a belief amounts to here. Since the goal of deliberation is to act on a correct belief about the right thing to do, conscience demands that I be as certain as I can be about the truth of my belief. This point would be unproblematic except that conscience as described in this passage is not supposed to be a faculty for *determining* moral truth, but rather for establishing the truthfulness or sincerity of my judgment of the case. Conscience is directed at my process of judgment and my grounds for thinking that I have done the best that I possibly can in arriving at my judgment. Kant describes this certainty as a 'postulate' because he remains skeptical that an agent can actually achieve transparency to self in the process of moral judgment, most notably because of the opacity of the actual motivations of the will.

Kant's wariness about the authority of conscience is also evident in his hesitant phrasing: conscience 'could be defined' as '*the moral faculty of judgment, passing judgment upon itself*', but Kant is not yet ready to actually define it as such. The 'prior clarification of the concepts' that he thinks would be necessary proves to be highly elusive indeed, as my later discussion will show. We have already seen that Kant wants to secure the objectivity of moral obligation by making rightness or wrongness solely a matter of the understanding. In this passage he also distinguishes the understanding from practical reason, which is the act of bringing cases under laws. My reason as subjectively practical tells me (i.e. judges) whether an action falls under the law (which secures rightness or wrongness through the understanding), and then I judge the process of reasoning to tell if it is sound. We thus have a two-step process of deliberation. As with all higher-order theories, the question arises of whether or not judgment is thereby caught in a regress. Who is to say, after all, whether or not my conscience itself is sound? While Kant does not explore this issue in the *Religion*, it will be central in the subsequent passages on conscience. Another pressing problem in determining the authority of conscience is that as such a final examination of the process of judgment, it needs to have 'at its disposal' all of the resources of deliberation, resources that then become difficult to distinguish from conscience itself.²⁶

²⁵ Cf. Hoffman, 'Gewissen als Praktische Apperzeption', p. 425.

²⁶ In his parenthetical remark on casuistry and 'the case of conscience', Kant both suggests that conscience is the site of adjudicating moral conflict and casts suspicion on the ability of conscience

At the end of the long passage cited above Kant makes use of the courtroom metaphor that he will expand in the *Metaphysics of Morals*. The second-order judgment of conscience addresses the question of whether a full examination of the case of action has taken place. To reach this judgment, the person calls himself as a witness. I am the person with the best access to my own deliberative process, so I am best qualified to serve as a witness to what did or did not take place in reaching this decision. In the tradition of theorizing conscience, this type of ‘inner witness’ view grants conscience the least authority, for it does not give conscience a role in actually determining the rightness of action.²⁷ But Kant’s own language here, identifying conscience with second-order moral *judgment*, is in tension with the inner witness account. In the imagined courtroom, the witness presumably lays out what did or did not take place in the first-order judgment. As a second-order judgment, conscience must also double-check, as it were, the results of the first-order judgment. If so, is there really any distinction here between the first and second order? Kant’s idea seems to be that my conscience does repeat the first-order judgment, but in doing so I *impute* the action to myself as a moral agent. The failure of this self-imputation is a lack of certainty about the rightness of the action, and I am warned not to act. If conscience can contravene the first-order judgment, however, we should wonder whether its authority really can be restricted.

3. A Limited Role for Conscience?

Thomas Hill has presented an elegant interpretation of conscience in Kant that gives conscience an important, but ‘limited role’ in moral reasoning.²⁸

to do so. By invoking a dialectic of conscience here, Kant implies that in trying to adjudicate how cases fall under laws, conscience oversteps its proper bounds. But it is not clear within cases of moral conflict just what the bounds of conscience are. We find in the lecture notes on ethics from the year after *Religion* an elaboration of the ‘*casus conscientiae*’ that is ambiguous about conscience’s role in deliberation. After favorable comments on establishing ‘a system of casuistics in morals’, the lecture notes refer to conscience as having ‘to do, not with duties themselves, that we seek to determine, but with *adiaphora*, that are made analogous to duty...refraining from all work on Sundays, fast-days etc. In such usages we fabricate a morality’. *Kant.Ak.* 27:619-20; *Lectures on Ethics*, ed. Peter Heath and J.B. Schneewind, trans. Peter Heath (Cambridge: Cambridge University Press, 1997), p. 362.

²⁷ See C.S. Lewis, *Studies in Words* (Cambridge: Cambridge University Press, 1967), pp. 181-213.

²⁸ Thomas Hill, ‘Four Conceptions of Conscience’, in Hill, *Human Welfare and Moral Worth: Kantian Perspectives* (Oxford: Clarendon Press, 2002). Though Hill’s essay works towards a Kantian view of ethics, and thus cannot be read simply as exegetical, I take it that on the central points of his discussions of the fourth conception of conscience, namely Kant’s, he is simultaneously doing exegesis and constructing the most plausible role for conscience in contemporary ethics.

In this section I present Hill's argument and consider possible alternative conceptions of conscience's authority. In what I am calling the *Limited Role (LR)* thesis, Hill identifies two accounts of conscience in Kant. First, Hill formulates Kantian conscience in terms of an affective awareness linked to the concept of imputability. He writes that conscience for Kant 'imposes a painful awareness of two distinguishable things: (1) that what we have done (or intend to do) is at odds with what, even in our own judgment, is [right] in the circumstances and (2) that the act is fully imputable to ourselves as a free agent'.²⁹ Hill brings the affective side of conscience to the fore, aiming to preserve our ordinary sense of conscience's immediacy against the view of conscience as a rational 'supervisory capacity' that he associates with Butler. Conscience is an affectively laden consciousness that my action is in conflict with my judgment, and that I cannot blame this conflict on something external to myself. That the 'act is fully imputable' means that I bear full responsibility for it, no matter the natural inclinations that I claim 'led me to do it'. Hill draws the second account of Kantian conscience primarily from the *Religion* text. He writes that it is a 'special, but quite broad, duty of due care',³⁰ a duty to appraise our actions with all due diligence. Conscience involves an assessment of our deliberative effort, of our moral attentiveness and perspicuity in arriving at a moral judgment.

Putting these two accounts of conscience together, Hill writes that 'conscience can acquit or condemn with regard to accusations of both violations of first-order duties (e.g., truth telling) and failures to fulfill the second-order duty of due care in scrutinizing and appraising our acts diligently (by "holding them up" to our judgment of the first-order duties)'.³¹ As a second-order judgment, conscience can raise accusations concerning the first-order judgment or the second-order duty to appraise diligently the first-order duty. The main limiting element in this account is that conscience is only active *after* a first-order moral judgment. Hill thus claims that Kant separates the activity of conscience 'from the general activities of moral deliberation, reasoning, and judgment'.³² The idea is that the authority of all of these capacities is independent of conscience, and that conscience is only brought to bear once these three activities have already run their course. Conscience can 'come into play only after one has made, or accepted, a moral judgment'.³³

²⁹ Hill, 'Four Conceptions of Conscience', pp. 301-302.

³⁰ Hill, 'Four Conceptions of Conscience', p. 302.

³¹ Hill, 'Four Conceptions of Conscience', p. 303.

³² Hill, 'Four Conceptions of Conscience', p. 301.

³³ Hill, 'Four Conceptions of Conscience', p. 299.

Hill thinks that Kant does justice to the ordinary usage of ‘conscience’, for conscience on this view is an instinct that is able to arouse powerful feelings. In so far as conscience *presupposes* the other faculties involved in practical reason, however, it is not a ‘mere felt clue’. Conscience scrutinizes the workings of the other faculties in moral action and arouses a painful feeling when my actual willing runs contrary to what I know to be my duty. Hill writes that the apparatus of practical reason is ‘the essential background assumption that enables us to think of conscience and conscientious judgment as having motivating force’.³⁴ The motivating force of conscience is conditioned by the representation of the moral law, so conscience is not what Kant calls a pathological feeling. Just how conscience as a motivational force is supposed to function in moral deliberation is left rather up in the air in Hill’s account. I will argue later (in section 6) that this motivational role for conscience implies that it is more fundamental than Hill claims. I will only note for now that given the importance of moral motivation in Kant’s account of moral worth, there is reason to think that conscience would have a more central role than Hill ascribes to it.

My main concern with Hill’s thesis has to do not with motivation, but with judgment. To understand the issues here, we need to examine more closely the language of ‘presupposition’ that Hill uses to cordon off conscience from the more familiar aspects of Kantian morality. He writes: ‘That someone has a conscience is a presupposition of his or her being a moral agent’,³⁵ and ‘Kant’s conception does not deny, but in fact presupposes, the possibility of objective moral judgments’.³⁶ Hill’s two presupposition claims run in opposite directions: conscience is a condition that makes moral agency possible, and conscience is conditioned by the prior operation of moral capacities. So first, moral agency presupposes conscience: if there is no conscience, then there is no moral agency. Second, conscience presupposes quite a lot about the moral universe and our access thereto (via the other, supposedly distinct faculties): if there are no objective moral judgments and well-functioning faculties, then there is no possibility of a well-functioning conscience.

These presupposition claims are not as ‘limiting’ as they initially appear. In the claim that moral objectivity is presupposed by conscience, the meaning of the claim of ‘objectivity’ for moral judgment is unclear. While the idea that conscience is distinct from the source of moral principles (what Kant calls the practical understanding), it is much less clear how moral deliberation and judgment in specific cases can be kept separate from conscience. Hill himself

³⁴ Hill, ‘Four Conceptions of Conscience’, p. 299, n. 40.

³⁵ Hill, ‘Four Conceptions of Conscience’, p. 301.

³⁶ Hill, ‘Four Conceptions of Conscience’, p. 307.

worries in another essay about the idea of an ‘objectively right’ judgment, which ‘raises questions’ in a Kantian framework, for in such a framework the perspective of the agent is essential. Hill writes, ‘Unlike many consequentialists, Kant is not primarily concerned with what is “right” independently of the knowledge and understanding of the agent’.³⁷ The best one can do for the objectivity of judgment is ‘what reason would prescribe *given a correct assessment of the facts, a clear understanding of the basic moral law, and no distorting influences on judgment*’.³⁸ I think that Hill’s remarks here on the objectivity of specific moral judgments in Kant are on target. The problem is that the third of the italicized desiderata is precisely what conscience itself is supposed to guarantee. If that is the case, then it does not make sense to say that objective moral judgment is presupposed by conscience.

The other presupposition claim, that moral agency presupposes conscience, admits of weak and strong interpretations. I take it that Hill’s weak interpretation of this claim is intimately related to his point that conscience is the awareness that an action is ‘fully imputable’ to me. In the many passages in which Kant writes of conscience as imputing an action to myself, what is at issue is imputation by a free rational being, not legal imputation by other agents. In the passage in the *Critique of Practical Reason*, Kant writes that conscience unavoidably holds my actions before me as products of my free noumenal causality. The timeless judgments of conscience thus attest to the noumenal character of my agency as a free being. The weak claim that Hill endorses is that conscience is simply the awareness or consciousness that I am responsible for my moral judgments and for the agreement of my actions with those judgments. I call this the *Post-Judgment Awareness View* of conscience because this account obviously works best in the retrospective mode: I have already judged and acted, and conscience is the awareness that I cannot make any excuses for what I have done, and I cannot help feeling bad if I have acted against my best judgment. This is part of what it means to be a moral agent, though it is a relatively minor or limited part.

There are two stronger senses in which conscience and imputation could be intertwined, both of which would cast doubt on the *LR* thesis.³⁹ First,

³⁷ Thomas Hill, ‘Punishment, Conscience, and Moral Worth’, in Hill, *Human Welfare and Moral Worth: Kantian Perspectives* (Oxford: Clarendon Press, 2002), p. 349, n. 26.

³⁸ Hill, ‘Punishment, Conscience, and Moral Worth’, p. 349, n. 26.

³⁹ Hoffman, ‘Gewissen als Praktische Apperzeption’, comes close to acknowledging the tensions that conscience creates within Kant’s overall view, but Hoffman’s concern to present a coherent and unified picture of Kantian conscience prevents him from developing the more radical elements in Kant’s writings. Thus Hoffman accepts Kant’s claim from the *Religion* that conscience can be restricted to merely formal second-order testing (he seems to accept something like Hill’s *LR* thesis), but he also makes the following more substantive claims. He writes that

conscience could be a self-consciousness *that makes the imputation of deeds possible*. One role that Kant ascribes to conscience is self-consciousness of the will. As such a capacity, conscience can be seen as enabling a subjective unity of the diverse aspects of a case of action, and thus as enabling the deed to be imputed to a unitary subject. I call this the *Self-Consciousness in Deliberation View*, which contrasts most obviously with the Post-Judgment Awareness View in that it accompanies the entire process of first-order deliberation. The second strong conceptual connection of conscience and imputation would hold that conscience simply *is* the judgment that imputes actions to ourselves at the close of deliberation. Conscience on this view is the all-things-considered judgment that this specific action is the one that I should perform in this situation. All the facts about the case and everything relevant to the case in my stock of moral beliefs will have been brought into deliberation, and the role of conscience here is to judge the unity of those elements in a representation of an action as my duty. I call this the *Unifying Judgment View*. This is the most natural way to read Kant's claims about conscience as rendering judgment. We can even read this idea out of the *Religion* passage, for the demanded certainty about the action 'I am to perform' makes sense as a demand that I impute the action to myself, prospectively. I must take up a judgment in the abstract, so to speak, and make it a judgment *for me*, an intention that I am willing to avow as my own. Notice that on this view the work done by Hill's weak view will be included within the judgment. But since on this view conscience actually makes the final judgment, rather than simply seconding a previous judgment, the role for conscience here would be greatly expanded. In the remainder of this paper I will be examining the evidence that Kant holds the two stronger claims about conscience and imputation. These two claims are compatible with each other, but they are very different, and the tensions between them lead to some of the more perplexing and provocative aspects of Kant's account.

4. The Courtroom of Conscience

In contrast to the cautious description of conscience in the *Religion*, in the *Metaphysics of Morals*, under the heading 'On Man's Duty to Himself as His

conscience in Kant is meant to address 'the question of a simultaneous intuitability and intelligibility of our practical *Dasein* in actions' (p. 437), 'the self-perception of the rational and free being as such in action' (p. 440), and the 'self-constitution of practical intelligibility in actions' (p. 442). I do not have a problem with these formulations as indicative of the tendency of Kant's late thinking on conscience, but they are certainly not compatible with the LR thesis, and they stand in tension with some of Kant's core commitments about the limits of self-knowledge.

Own Innate Judge’,⁴⁰ Kant gives a bold account of the courtroom of conscience. Though Hill cites this passage in support of the *LR* thesis, it is far from clear that Kant is talking about an activity that takes place *after* the first-order judgment of the case. Here is Kant’s set-up of the metaphor:

Every concept of duty involves objective constraint through a law (a moral imperative limiting our freedom) and belongs to practical understanding, which provides a rule. But the internal *imputation* of a *deed*, as a case falling under a law (*in meritum aut demeritum*), belongs to the *faculty of judgment (iudicium)*, which, as the subjective principle of imputing an action, judges with rightful force whether the action as a deed (an action coming under a law) has occurred or not. Upon it follows the conclusion of *reason* (the verdict), that is, the connecting of the rightful result with the action (condemnation or acquittal). All of this takes place before a *tribunal (coram iudicio)*, which, as a moral person giving effect to the law, is called a *court (forum)*. – Consciousness of an *internal court* in the human being (‘before which his thoughts accuse or excuse one another’) is *conscience*.⁴¹

Kant thus contrasts the objective constraint through a law with the subjective principle of imputing an action. But the relation of this ‘faculty of judgment’ to conscience is not immediately clear from this passage, which can be interpreted as supporting any of the three models of conscience and imputation I sketched at the end of the last section. These readings are all open because the claims that ‘All of this takes place before a *tribunal*’ and that conscience is the consciousness ‘of an internal court’ are ambiguous. Kant could be saying that the law, judgment and reason are the actors within the tribunal, and conscience is just the consciousness of the imputation and verdict that has already taken place (it is the mere consciousness *that* they have taken place). It is therefore possible to read this passage along with Hill as supporting the Post-Judgment Awareness View. This is an awareness of the correspondence or lack thereof between the subject’s intention/action and the first-order objective judgment of what is right in a given situation.

But there are better reasons to read this ‘consciousness of an internal court’ as a self-consciousness in the process of judgment that makes an action a possible action for me. This corresponds to the Self-Consciousness in Deliberation View of conscience and imputation. Taking conscience as the consciousness of the court of moral judgment, we can conceive of these faculties as separable, but they are considered as a unity under the auspices of conscience. Conscience on this view would be a self-consciousness that makes the imputation of actions possible in the first place by demanding transparency from the tribunal’s

⁴⁰ *Kant.Ak.* 6:437-40; *Practical Philosophy*, pp. 559-62.

⁴¹ *Kant.Ak.* 6:437-38; *Practical Philosophy*, pp. 559-60.

different actors. Conscience makes sure that all evidence is available, for it is the self-consciousness of everything that takes place within judgment.

Yet there are also good reasons to think that Kant is contrasting with ‘objective constraint’ the operation of conscience as the all-things-considered judgment in which the agent imputes the action to herself. This would support the Unifying Judgment View. After all, how else should we read the phrase ‘as a moral person giving effect to the law’? On this conception I judge whether a proposed action really is qualified to be *my* action. The reference to the ‘subjective principle of imputing an action’ seems to remove us completely from the first-order judgment about a case in the world, but we need to remember that maxims of action just are subjective principles of action.⁴² The success or failure of imputing the maxim through the law is therefore hard to separate from the judgment of whether or not the action is right. In fact, in the rest of the description of conscience Kant invests conscience with so much authority that the first-order/second-order distinction becomes completely untenable.

The Unifying Judgment View of conscience and imputation comes out in Kant’s elaboration on conscience as the subjective site of the central conundrum of autonomy. Kant claims that in theorizing conscience we can answer the question, how can I be both the author of the law and bound by the law? Kant writes that this ‘original intellectual and (since it is the thought of duty) moral predisposition’ is ‘peculiar’ because although it seems to be an internal activity of a single individual, the individual must take himself to be divided or doubled. In this internal courtroom the defendant and the judge cannot be the same person, for then the defendant would always win, the prosecutor would always lose. Conscience must be figured as someone other than the person acting or else ‘conscience would be in contradiction with itself’.⁴³ In conscience the agent has ‘to think of *someone other* than himself as the judge of his action’.⁴⁴ The subject must stand (‘trembling’) before a court that is on the one hand entrusted to the subject, but which on the other hand has its own innate authority. Kant writes,

⁴² Hill could support his account with the comparative language Kant uses here. The sentence in which Kant writes of whether the action ‘has occurred or not’ implies a comparison between what I am about to do with what I have judged to be right (‘an action coming under a law’). So too, ‘the connecting of the rightful result with the action’ seems to refer to a judgment (‘the rightful result’) and the action I am about to undertake. My claim is that even if Kant wanted to keep the objective judgment and subjective imputation separate, the rest of the account shows that he is not entitled to such a separation.

⁴³ *Kant.Ak. 6:438; Practical Philosophy*, p. 560.

⁴⁴ *Ibid.*

the human being as the subject of the moral lawgiving which proceeds from the concept of freedom and in which he is subject to a law that he gives himself (*homo noumenon*) is to be regarded as another (*specie diversus*) from the human being as a sensible being endowed with reason, though only in a practical respect.⁴⁵

The ideal self is the lawgiver and the real self is the finite being attempting to live up to that law. The law-giver sets the terms of moral justification, and the real self tries to judge what satisfies those terms and to determine the will such that it is motivated by those same reasons.

Kant does write that this second self could be an actual person, but he thinks that reason requires an ideal self that can only be conceived in terms of divinity. The judging and prosecuting conscience must be thought of as an omniscient, all-commanding, and omnipotent moral being, namely God.⁴⁶ The idea of an omniscient and omnipotent God secures the notion of conscience's efficacy, what Kant calls the 'power...to give effect to his laws'. In conscience the agent is supposed to know all the circumstances, of her own experience and of the case in the world, and to be able to motivate the act through the law itself. Kant calls this God 'the authorized judge of conscience' and 'a scrutinizer of hearts', and claims that conscience is 'the inner judge of all free actions'.⁴⁷ This discussion gives very strong support to both the Self-Consciousness in Deliberation View and to the Unifying Judgment View of conscience. In fact, this passage runs those two views together, since here Kant takes conscience to be an actual judgment of the integrity of my assessment of the case that presupposes self-conscious access to everything I know about myself and about the circumstances.

Given this full-blown account of conscience as the divine inner judge, it is very hard to see how the functions of conscience for Kant can still be kept separate from first-order judgment and deliberation. Conscience, as Kant describes it, has all the resources of practical reason, which means that the conclusion of conscience is *indistinguishable* from the conclusion of deliberation. The question for deliberation is to determine what my maxim of action will be. The question before the court is the subjective question, 'Do you *believe* that the action you are about to perform is the right (moral) action?' If you do not believe it is, then you are warned by your conscience. But how do you assess the soundness of your belief? By the same process that would answer

⁴⁵ *Kant. Ak. 6:438; Practical Philosophy*, p. 560.

⁴⁶ Of course God thus conceived is only a subjective, regulative ideal, and no grounds for deducing God's objective existence can be determined from this experience of ourselves. For the theological background of this view, see Heinz Kittsteiner, *Die Entstehung des modernen Gewissens* (Frankfurt am Main: Insel, 1995), pp. 274-83.

⁴⁷ *Kant.Ak. 6:439; Practical Philosophy*, p. 561.

the simple question, ‘What *is* the right (moral) action?’, namely by considering all the appropriate evidence and arriving at a judgment about the most rational/moral action in these circumstances. The separation of a first-order judgment and a subsequent second-order judgment judging itself begins to seem like a rather desperate assertion of the ideal objectivity of moral judgment. In practice it is a distinction without any real difference. In terms of the central instability in autonomy, the distinction is supposed to maintain the objective law’s priority over the authority of the individual self. But since conscience gives the final word on the rightness of an action, and gives my intention the seal of certainty, in cases of moral complexity the judgment of the case is better ascribed to the overall, comprehensive activity of conscience than to faculties that one can conceive of operating independently of this or that subject. We can thus say that conscience is the judgment whereby an action is imputed to the subject, and that this judgment just is the judgment of the case of action.

5. Conscience as Apperception

There is a major problem confronting the account I just gave of conscience as a comprehensive judgment of an action. In several places, including within the *Metaphysics of Morals* itself, Kant claims that an ‘erring conscience’ would be an absurdity. Yet it is clear, as Kant acknowledges, that our judgment of cases in the world is subject to error. If conscience’s activity is supposed to be incompatible with error, there is a great deal of pressure to think of Kant’s view of conscience in a minimal sense (along the lines of the Post-Judgment Awareness View). One might suppose that there is an obvious way around this problem. Following the above analogy one could say that since the idea of God, the real inner authority behind conscience, is incompatible with the idea of error, if I do make a mistake in my judgment, then it was not an authentic judgment of conscience (even though I thought I was acting according to my conscience). Though this is not what Kant means in denying an erring conscience, there is something in his account that pushes him towards this view. As we just saw, Kant tends to conflate the Self-Consciousness in Deliberation View and the Unifying Judgment View of conscience, to conflate conscience as the moral self-consciousness that enables imputation and conscience as the judgment of imputation itself. So in his discussions of conscience as God he writes that the idea of accountability to this lawgiving God is ‘contained in every moral self-consciousness’. If the idea of a lawgiving God just defines what it is to be morally self-conscious, then I only am an authentic moral agent when I have judged correctly. This view raises problems with the imputability of mistaken judgments, as we shall see at the end of this section.

But first we need to examine the claims in Kant's other discussion of conscience in the *Metaphysics of Morals*, which give us reason to keep separate, within an overall account of conscience, the Self-Consciousness in Deliberation and the Unifying Judgment Views. In this section I examine the main text that denies the idea of an erring conscience, and I present an interpretation of conscience as moral self-consciousness. I criticize the interpretations of the no-erring-conscience claim by Hill and Allen Wood, arguing that they explain it by explaining it away. I also introduce considerations to suggest that Kant himself was not in control of the various roles of conscience, and that he is led to a claim about the *sufficiency* of conscience for morally guiltless action that brings out the instability in his conception of autonomy.

In the Introduction to the 'Doctrine of Virtue', conscience appears as one of four concepts (along with 'moral feeling', 'love', and 'respect') under the heading 'Concepts of What is Presupposed on the Part of Feeling by the Mind's Receptivity to Concepts of Duty as such [*Ästhetische Vorbegriffe der Empfänglichkeit des Gemüts für Pflichtbegriffe überhaupt*]'.⁴⁸ These are 'moral endowments' that humans must possess if they are to act on duty at all. Though it is listed alongside other necessary 'aesthetic' preconditions for morality, conscience is not a distinctive kind of feeling (as the other three are). Kant presents conscience as both an 'unavoidable fact' and as a deliberate activity.⁴⁹ He denies that we can speak intelligibly of a moral agent who has no conscience, writing that the idea of a duty to have a conscience would require an agent to start without a conscience, but 'if he really had no conscience, he could not even conceive of the duty to have one, since he would neither impute anything to himself as conforming with duty nor reproach himself with anything as contrary to duty'.⁵⁰ Conscience is necessary to 'even conceive of' a duty, which I take to mean that one could not even have a moral will without the self-consciousness through which deeds become imputable.⁵¹

This passage is so difficult in large part because Kant does not distinguish this self-consciousness from conscience as delivering judgments that actually do impute an action to the subject. That is, he does not distinguish between the Self-Consciousness in Deliberation View and the Unifying Judgment View. He writes that I can 'pay no heed to its judgment', but not having such

⁴⁸ *Kant.Ak*, 6:399; *Practical Philosophy*, p. 528.

⁴⁹ *Kant.Ak* 6:400; *Practical Philosophy*, p. 529.

⁵⁰ *Kant.Ak* 6:400–401; *Practical Philosophy*, p. 529.

⁵¹ The importance of conscience is supported by a parenthetical remark a few pages later in the Introduction to the 'Doctrine of Virtue', where Kant writes 'conscience was treated earlier, as the condition of all duties as such'. *Kant.Ak* 6:407; *Practical Philosophy*, p. 534.

a judgment is inconceivable. Kant emphasizes in this passage that conscience is ‘directed...to the subject’, for it concerns the imputation of the action to the agent as conforming or not conforming with a duty. Kant then claims that ‘an erring conscience is an absurdity’, and unpacks that claim by writing,

For while I can indeed be mistaken at times in my objective judgment as to whether something is a duty or not, I cannot be mistaken in my subjective judgment as to whether I have submitted it to my practical reason (here in its role as judge) for such a judgment; for if I could be mistaken in that, I would have made no practical judgment at all, and in that case there would be neither truth nor error.⁵²

The subjective judgment of conscience is necessary for there to be moral judgment at all. The key to unpacking this passage is not to take the ‘subjective judgment’ as an actual judgment, but rather to understand it as the self-consciousness that makes possible judgments imputing an action to the subject. ‘Whether I have submitted it to my practical reason’ is not something I explicitly judge, but rather something that conscience as self-consciousness makes possible. Kant’s no-erring-conscience claim is best understood as restricted to the claim that there is a basic act of self-consciousness that is constitutive of a judgment of duty. That is just what it means to say that *there is no practical judgment* without an accompanying act of self-consciousness.

I read this self-consciousness as a version of the well-known claim from the first *Critique* about the ‘I think’ that ‘must be able to accompany all my representations’.⁵³ If I am to properly represent to myself an action as my duty, I must be conscious of myself in all the representations relevant to the action. Because this self-consciousness must be *practical relevant*, it should be conceived as ‘I think I am willing’. It is a presupposition of practical judgment that I am aware of everything relevant to my willing in this case, and that I am *taking* it to be relevant to my willing. We have seen some evidence for this reading already in the *Religion* passage. Another clue in support of this reading comes in a passage from the *Vigilantius* lectures in which Kant claims that conscience is the practical form of apperception. He writes,

The doctrine of conscience is of the greatest importance in morals. *Conscientia*, taken generally, is the consciousness of our self, like *apperceptio*; *in specie* it involves consciousness of my will, my disposition to do right, or that the action be right, and thus equals a consciousness of what duty is, for itself.⁵⁴

⁵² *Kant.Ak* 6:401; *Practical Philosophy*, p. 529.

⁵³ *Kant.Ak*. B131-32; *Critique of Pure Reason*, ed. and trans. Paul Guyer and Allen W. Wood (Cambridge: Cambridge University Press, 1998), p. 246.

⁵⁴ *Kant.Ak*. 27:613-14; *Lectures on Ethics*, p. 357.

Kant's distinction here between conscience 'taken generally' and conscience 'in specie' is quite close to my distinction between the Self-Consciousness in Deliberation View and the Unifying Judgment View. The 'in specie' view is just the all-things-considered judgment of the case that I consciously impute to my will (notice how much Kant packs into the description). But the claim here about the 'consciousness of our self', the claim of the Self-Consciousness in Judgment View, describes a much thinner self-consciousness of the will in the process of deliberation. The relation of the two views is like the relation in the theoretical case between transcendental apperception and the conscious representation of an object. Without a possible unity of self-consciousness, the same 'I think' in all the elements of a representation at once, there can be no combination of the manifold that could count as the representation of an object. This condition is a basic requirement to conceive of normatively constrained representation of objectivity at all. In the practical case, this transcendental unity is conscience as a self-consciousness that makes the self-imputation of duty possible. This act guarantees that all elements in a practical case are accessible to me and can contribute to a motivationally efficacious representation of my duty.⁵⁵

How can conscience be both a bare act of self-consciousness *and* a judgment? Conscience thus conceived is complex, but also coherent as a single concept. Kant is taking conscience to be *both* a presupposed act of self-consciousness in the process of deliberation and the judgment of the *unity of self-consciousness* that closes deliberation. There is a basic act of self-consciousness that makes imputation *possible* in the first place, and there is an act of judging the unity of self-consciousness in a specific piece of moral deliberation. Much of the interpretive difficulty in providing a unified concept of conscience in Kant stems from his failure to keep these two senses separate. In the claim that an erring conscience is an absurdity, Kant is referring to the primitive act of self-consciousness. He admits that I can always be wrong about whether an action is in fact permissible or impermissible, whether from lack of knowledge of either moral principles or the circumstances at hand. The question is how to understand the constitutive character of conscience

⁵⁵ Though I cannot pursue it within the scope of this essay, this figure of conscience as self-consciousness could be read as the key to understanding Kant's enigmatic discussion of the categories of freedom in the *Critique of Practical Reason*. He writes that the categories 'subject a priori the manifold of *desires* to the unity of consciousness of a practical reason commanding in the moral law, or of a pure will'. *Kant.Ak.* 5:65; *Practical Philosophy*, p. 192. For an excellent account of the categories that takes this use of practical self-consciousness seriously (though not utilizing the later discussion of conscience), see Claudia Graband, 'Das Vermögen der Freiheit: Kants Kategorien der Praktischen Vernunft', *Kant-Studien* 96 (2005), pp. 41-65.

for first-order judgment without the infallibility of conscience (falsely) implying that objective judgment is infallible.

The claim that ‘an erring conscience is an absurdity’ is unintelligible without the distinction that I just introduced between a basic act of self-consciousness and the judgment of the unity of self-consciousness in deliberation. The difficulty with Kant’s tendency to conflate these ideas is on display in the two best treatments of this issue in the literature, by Hill and Allen Wood. Hill explains the claim that ‘I would have made no practical judgment’ without conscience by invoking as a ‘background fact’ the knowledge that I have submitted a case to my judgment. Hill claims that this ‘is not literally...what conscience judges’.⁵⁶ But what Hill calls a background fact is the basic activity of conscience as apperception that makes imputation of an action possible. Hill accounts for the no-erring conscience claim by arguing that conscience is only active in cases in which *I have in fact* submitted my action to moral standards (cases in which I have reached a first-order judgment). If I have not thus submitted my acts to the law, conscience ‘never operated and so cannot have yielded a false verdict’.⁵⁷ Hill needs this kind of claim in order to preserve the main tenet of his *LR* thesis, namely that conscience only operates *following* the first-order judgment. But Hill’s explanation raises problems for the coherence of his overall account. Recall that a key dimension of Hill’s reading of conscience takes it to be the painful awareness ‘that the act is fully imputable to ourselves as a free agent’.⁵⁸ Combining this claim with the claim that conscience is only active when I have already held up the case to moral standards, the result is the claim that if I do not compare my action to moral standards, then I have no awareness that the action is fully imputable to me. But on Hill’s own account, conscience makes itself felt as a ‘painful awareness’ that all my actions are imputable to me. Hill’s difficulty arises because he does not separate Kant’s claims for conscience as a constitutive presupposition and as a deliberative activity (as we have seen, Kant himself tends to conflate these two dimensions).

Allen Wood takes up the issue by considering whether Kant’s no-erring-conscience claim denies the possibility of ‘a mistaken judgment concerning whether I have held my actions up to the rational standards of moral judgment’.⁵⁹ As Wood goes on to point out, Kant himself insists that we often delude ourselves (i.e., make mistakes) in thinking that we have held up our actions to rational standards when we have not. So if Kant were denying error

⁵⁶ Hill, ‘Four Conceptions of Conscience’, p. 303, n. 50.

⁵⁷ *Ibid.*

⁵⁸ Hill, ‘Four Conceptions of Conscience’, p. 302.

⁵⁹ Wood, *Kantian Ethics*, p. 190.

in the judgment of ‘whether I have held my actions up to the rational standards’, he would seem to be inconsistent with his own observations. Wood thus writes, ‘Kant’s argument is this: In the self-deceptive belief that I have acted conscientiously when I have not, there has been no genuine judgment of conscience at all, so there cannot have been an erroneous one either’.⁶⁰ Wood is saying that if a judgment of conscience is genuine, then I will know it as such (have an accurate belief), and that conscience is not genuinely operative (and thus cannot be erroneous) if the belief in acting conscientiously is self-deceptive. This puts all the weight of truth and error on the belief about acting conscientiously, but this first-person belief just is the claim of conscience as apperception that Kant is denying can be in error. By reading this belief as separate from the “genuine judgement of conscience,” Wood is deflecting the question of the erring conscience rather than addressing it.

But Wood’s distinction does partly overlap with my distinction between the Self-Consciousness in Judgment View and the Unifying Judgment View. What Wood considers a ‘genuine’ judgment of conscience is the judgment of the unity of self-consciousness at the close of deliberation. Kant thinks that such a judgment is always available to the agent, so that in cases of self-deception the agent just does not make explicit that demand for unity. The impossibility of error is the claim that apperceptive conscience is constitutive of first-order judgment, so that there could be no first-order judgment of duty without the condition under which the action is imputable to me as a duty. The question that remains is whether the idea of error is compatible with the judgment of unity at the close of deliberation, and what the consequences are if error in that judgment is ruled out.

My claim that conscience as self-consciousness is constitutive of moral judgment is also supported by Kant’s concern about a possible infinite regress in self-assessment. If conscience were an ordinary second-order judgment, Kant would have to find a way to *block the regress* in judgment that threatens theories of higher-order activity. By making conscience constitutive of judgment, he instead *prevents the regress from getting started*. Recall that in the *Religion*, Kant had called conscience ‘judgment passing judgment on itself’, which seemed to raise the threat of needing a judge for the judge, ad infinitum. In the passage from the Introduction to the ‘Doctrine of Virtue’, he writes, ‘To act in accordance with conscience cannot itself be a duty, for if it were, there would have to be yet a second conscience *in order for one to become conscientious* of the act of the first’.⁶¹ As moral self-consciousness, conscience is

⁶⁰ Wood, *Kantian Ethics*, p. 192.

⁶¹ *Kant.Ak.* 6:401; *Practical Philosophy*, p. 529, my italics.

how the agent becomes conscious of his duty, which I read as making imputation of a duty possible. Having a duty to act in accordance with conscience would require that I be able to impute the activity of conscience to myself. But that would mean having another conscience to make possible that self-imputation. If, on the other hand, conscience is constitutive of moral judgment, no regress can get started because as self-consciousness, conscience accompanies the process of deliberation rather than standing above it.

But Kant was not clear enough about the distinction of the constitutive self-consciousness of conscience and the unifying judgment of conscience that closes deliberation. He is therefore led to a conclusion about moral culpability that stands in tension with his insistence on the objective bindingness of the universal law and highlights the central instability in Kant's conception of autonomy. This tension comes to the fore towards the end of this discussion of conscience, where Kant claims, 'But if someone is conscious 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'.⁶² If I cannot, despite my best efforts, make a better moral judgment, you cannot blame me for not having done otherwise. So while the no-erring-conscience thesis does not imply that I can never be mistaken in my judgment about particular cases, it does imply that I cannot be considered guilty for those mistakes. This result follows from the idea that conscience *defines the limits* of what can be imputed to me as a moral agent. This means that conscientious mistakes of judgment cannot be morally imputed to me. Kant thus avoids facing the question of whether conscience as unifying judgment can err by changing the topic from truth and falsity to innocence and guilt.

Needless to say, this conclusion is surprising given his commitment to an impersonal objective criterion for duty in the moral law. Whereas conscience seems in other passages to be the site of guilt and remorse, in this discussion it seems to be the source of exculpations, for getting me off of the hook for my mistakes ('Don't blame me, I acted in accordance with my conscience!'). This odd conclusion reflects what I have identified as a central instability in Kant's concept of autonomy. My action counts as autonomously willed if I can impute my action to myself through the universal law of my reason. Kant tried in the 'inner God' passage to identify this act of self-imputation with the universal law itself. But in the no-erring-conscience claim and the no-guilt-if-I-have-followed-my-conscience claim, the constitutive nature of the pure self-consciousness of conscience stands in tension with the objective authority of the law. In the Fichtean terms with which I began this paper, the dynamics

⁶² *Kant.Ak.* 6:401; *Practical Philosophy*, p. 529.

of conscience have pushed Kant's moral theory towards a priority of the self to the law. Because of Kant's almost exclusive focus on the law as the basis of objective constraint, once the self does come to the fore he has no way to secure the subject's accountability for his conscientious mistakes.

6. Conscience and Motivation

In the preceding two sections I have considered the function of conscience in judgment solely along its cognitive dimension. In this section I turn to the role that conscience plays in *moral motivation*.⁶³ In the claim that conscience is like apperception, Kant referred to conscience as 'consciousness of my will, my disposition to do right'. I also noted in the last section that Kant included conscience among his list of 'aesthetic' concepts. But it is unclear how Kant can endorse a consciousness of the motivations of the will without violating his restriction on knowledge of the will in action. While Hill and Wood both assign to conscience an important motivational role, neither of them works out its implications for the standard picture of Kant's moral psychology. Kant assigns conscience a role in positively motivating the agent to action, rather than simply warning the agent away from action or castigating the individual after action has taken place.⁶⁴ This positive motivational role for conscience is very hard to make out from Kant's writings, but drawing several aspects of his

⁶³ This reading is directly at odds with that of H.J. Paton, who in his posthumously published 'Kant and Conscience', *Kant-Studien* 70 (1979), vol. 3, pp. 239-51, puts forward another limited role thesis. Paton tries to render Kant's view coherent and consistent by claiming that the purview of conscience is only the legality of actions, not the moral worth or goodness of the actions; he holds that conscience would conflict with central Kantian theses were it a figure of moral motivation. There is something very plausible about Paton's thesis, yet there are many passages in which Kant gives conscience a motivational role. What makes the claim *prima facie* plausible is that the question of imputation that is central to all of Kant's discussions of conscience is naturally a legal question. Indeed, most discussions of imputation in Kant focus on the question of how we can hold bad people responsible given that Kant thinks they are not fully free in acting badly. On my view the question of imputation is much broader than this; in part the stress on *self*-imputation determines the breadth of the problem, since no one is better suited to assess my *motivations* than myself, and it is hard to see how motivation could be bracketed in my deliberative self-assessment.

⁶⁴ Lehmann, *Kants Tugenden*, p. 57, acknowledges the tension between Kant's law-governed rationalism and his claims about conscience as affective when he writes: 'Our theme was the analysis of conscience in Kant. And there one can say that he did attend to the "phenomenal horizon" of conscience, even if his objectivism again and again forced him to look away in the formation of ethical concepts from conscience as experience. Instead of saying that Kant's rationalism bears irrational elements, one can say that he is a rationalist with a bad conscience'.

account together (and once again following Fichte's lead) I will suggest that we understand it as a feeling of psychic harmony or unity. This suggestion will lead us back to the instability in autonomy, and to the nature of the Fact of Reason.

In the courtroom of conscience passage in the *Metaphysics of Morals* Kant highlights the threatening pose that the agent's conscience can take against the agent herself. He writes, 'Every human being has a conscience and finds himself observed, threatened, and, in general, kept in awe (respect coupled with fear) by an internal judge'.⁶⁵ The judge 'pronounces the sentence of happiness or misery',⁶⁶ and it is presumably this misery following upon a wrong action that we are to fear from conscience. While Kant is describing a familiar aspect of conscience, there is reason to worry that he has made moral motivation too much a battle of affective forces struggling for superiority. In particular, the idea that there is an element of 'fear' in this self-relationship raises the concern that the motivation here will not be properly moral, even if it is related to the moral law's inner representative. We can put the worry in the terms given by Andrews Reath in his article on 'Kant's Theory of Moral Sensibility': it seems to support 'a model that would understand the moral motive to operate by exerting a force on the will'.⁶⁷ The worry with this awe-inspiring conscience is that what actually impels my will (to speak in this quasi-external way) to action is the fear that my inner judge will make me miserable.

Yet Kant's discussion of the inner courtroom also suggests an alternative model of how the motivational dimension of conscience informs my willing. The alternative to thinking of motives as 'exerting a force on the will' is thinking of *granting authority* to one sort of incentive or another. In the same *Metaphysics of Morals* passage, Kant calls the courtroom scene of conscience a 'matter of right [*Rechtssache*]' in which an inner conflict is adjudicated. We can thus see conscience as addressing *the right by which a motive is to determine my will*, a right that is naturally decided in a courtroom rather than on a 'battlefield' of motivational forces. As a defendant, I am always under suspicion of having acted on self-love, or planning to act on reasons secretly informed by self-love. Only if I act on the right reasons, with the right motivation, will I be acquitted of wrongdoing. I ask myself in the courtroom of conscience, have you given priority to moral considerations? Are there hidden prudential reasons at the heart of your action?

⁶⁵ *Kant.Ak* 6:438; *Practical Philosophy*, p. 560.

⁶⁶ *Kant.Ak*. 6:438; *Practical Philosophy*, p. 561.

⁶⁷ Andrews Reath, 'Kant's Theory of Moral Sensibility: Respect for the Moral Law and the Influence of Inclination', in Reath, *Agency and Autonomy in Kant's Moral Theory* (Oxford: Oxford University Press, 2006), pp. 8–32, p. 12.

Once we appreciate that these questions are part of deliberation itself, we have additional reasons for thinking, contra the *Limited Role* thesis, that the process of the inner courtroom is not merely a final stamp of approval on a previous judgment, or an affective warning against an action that contravenes one's judgment. The affective dimension of the courtroom shows that the process of deliberation and the process of determining one's motivation are inseparable. This point gives a more definite sense to Kant's claim that the inner ideal God of conscience 'must also *impose all obligation*, that is, he must be, or be thought as, a person in relation to whom all duties whatsoever are to be regarded as also his commands; for conscience is the inner judge of all free actions'.⁶⁸ The 'inner judge' might seem superfluous if it merely duplicates the work of the quasi-external objective judgment of the case. But we can now see that through conscience I relate the action to my moral disposition, so that my deliberation takes into account my motives for performing this or that action. This motivational dimension gives real bite to the idea that the self-imputation of an action is essential to the content of the judgment itself. For I can only know conclusively that my action is right if I achieve a certain transparency to myself about my motivations and their relevance to the case. Kant invokes the idea of God as a 'scrutinizer of hearts'⁶⁹ because such a judge is able to not only judge a case in the world, but also to give a thoroughgoing account of how my disposition informs and is informed by my judgment of the case.

Yet rather than take conscience as the inner threatening God to be Kant's last word on conscience as motivation, we should instead take the passage from the Introduction to the 'Doctrine of Virtue' as a more general and fundamental model of conscience's role in deliberation. Recall that Kant includes conscience among the 'Concepts of what is Presupposed on the Part of Feeling by the Mind's Receptivity to Concepts of Duty as Such', and distinguishes it from moral feeling, love and respect. He describes these concepts as '*subjective conditions of receptiveness to the concept of duty, antecedent predispositions on the side of feeling*'.⁷⁰ It is no easy task to figure out how conscience is supposed to meet this description, for unlike the other three concepts, conscience as described in this brief passage is not a feeling at all in any obvious sense. Kant claims that 'conscience is practical reason holding the human being's duty before him', yet he insists that conscience 'is not directed to an object but merely to the subject (*to affect moral feeling by its act*)'.⁷¹ Kant had just defined

⁶⁸ *Kant.Ak.* 6:439; *Practical Philosophy*, p. 561.

⁶⁹ *Ibid.*

⁷⁰ *Kant.Ak.* 6:399; *Practical Philosophy*, p. 528.

⁷¹ *Kant.Ak.* 6:400; *Practical Philosophy*, p. 529, my italics.

moral feeling as ‘a *susceptibility* on the part of free choice to be moved by pure practical reason (and its law)’.⁷² Moral feeling is the most general conception of a feeling of pleasure or displeasure that arises *from* the representation of my duty. Conscience is thus an *activity* of deliberation in which the agent herself *brings about* moral feeling through the representation of the case as her duty. We can therefore think of conscience as the *activity* through which I represent my duty to myself such that I produce moral feeling that motivates my action. As something I intentionally do, conscience stands for the actual judgment of the unity of self-consciousness at the close of deliberation. Conscience is also the basic activity of self-consciousness that must be presupposed for there to be receptivity to duty (this is conscience ‘taken generally’ rather than ‘in specie’).

We should take Kant’s doctrine of conscience, rather than the narrow doctrine of respect, as the key to Kant’s theory of ‘moral sensibility’. Only conscience is a general enough notion to accommodate all the dimensions of moral feeling, rather than simply the oppositional feeling of respect. This generality is obscured by Kant’s treatment of conscience as a fearful divinity inducing fear and respect. But it makes better overall sense of Kant’s various claims about conscience to see them all as aspects of the more basic idea of a capacity for self-imputation through which duty is represented in a motivationally efficacious manner.

What exactly does ‘affecting moral feeling by its act’ entail positive terms? Kant’s consistent position on the feeling that accompanies successful determination of the will is ‘self-contentment’. This is not a feeling of ‘enjoyment’, but rather ‘a negative satisfaction with one’s existence, in which one is conscious of needing nothing’.⁷³ This description is not terribly helpful. Shifting our attention back to the *Religion* discussion of conscience, we can ask whether this self-contentment is the same as the feeling of achieved *certainty* that an action I am about to perform is right. It is tempting to think of this certainty (which has the status of a postulate for Kant) as cognitive certainty plus the mere absence of a warning from conscience that the action is wrong. But this does not do justice to the idea that conscience is the consciousness of ‘my disposition to do right’. Conscience as moral self-consciousness should rather be taken to be the capacity to judge a determinate case of action in relation to one’s entire ‘motivational set’, to borrow a phrase from Bernard Williams (that admittedly may seem out of place in this context). Of course to remain Kantian, the leading aspect of this self-consciousness must be the demand to

⁷² *Kant.Ak.* 6:400; *Practical Philosophy*, p. 529.

⁷³ *Kant.Ak.* 5:117; *Practical Philosophy*, p. 234.

prioritize moral reasons over all prudential reasons. But this is perfectly compatible with moral emotions playing a central motivational role, as long as I am clear that those emotions have authority only by virtue of moral reasons (the mere presence of an emotion is not itself a reason). It is also possible that my judgment in a new case will lead me to revise that motivational set. Reading conscience in this way has the added benefit of making sense of what is involved in the imperative to ‘cultivate’ one’s conscience. To cultivate my conscience means to become more responsive to duty, in that when I confront a situation I judge it as a case of duty and I arouse in myself the moral feeling that motivates action.

7. Conclusion

I return in conclusion to the question of how conscience can be the basis of a more plausible Fact of Reason argument. I discussed in section 1 Proops’s view that Kant took the uniformity of conscience for all moral agents as evidence for the Fact of Reason. Kant’s idea of accountability to God in every moral self-consciousness does closely resemble the original *Factum*, and it suffers from the same problems. It is, for one thing, too culturally specific to be plausibly assumed to be the actual moral psychology of all human beings. But I have been building a case for thinking that there is a more basic – and much more defensible – activity underlying the explicit judgment of unity in conscience. My treatment of conscience as a basic act of self-consciousness is closer to Franks’s interpretation of the *Factum* as an act and to his identification of that act with the first-person standpoint. The limitation of Franks’s view is that it only licenses an inference to freedom for the person responding to the examples in Kant’s text. The argument for conscience as a basic self-consciousness constitutive of judgment, on the other hand, is a claim that no one could be a moral agent at all without such a conscience. This is an argument about the practical point of view, but one that can be defended without an appeal to the responsiveness of the individual reader.

I have claimed that the primacy of the basic act of practical self-consciousness reveals a priority of the self to the lawfulness of the understanding. The question is, can conscience thus conceived serve the two main functions of the *Factum*, awareness of the moral law as the principle of moral judgment and awareness that the moral law can in fact motivate, or determine the will? I have marshaled the evidence from Kant’s texts that conscience is essential to moral judgment and that it is involved in moral motivation. Conscience can thus do the work of the *Factum*, though this will mean lowering our expectations about the extent to which the moral law can provide determinate

judgments in the abstract. Demoting the form of lawfulness to a merely heuristic role, the moral law that defines autonomy should be taken to be a law of self-determination plain and simple. Or as Fichte puts it, ‘the law of absolute agreement with oneself’.⁷⁴ This can be read as a version of the Unifying Judgment View of conscience that demands unity of self rather than unity under the form of lawfulness. One could object to this idea that without the form of universality as *constitutive* of the moral law, conscience could not serve as the *Factum* because there is no principle of judgment. My all too brief response to this worry is that in cases of moral complexity the form of universality itself cannot serve as a principle of judgment, so it only *appears* to be better suited to determine judgment. Though conscience as unity of self-consciousness is not going to provide a rule-like procedure for arriving at moral judgment, as a principle of *integrity* it seems to me to be a better principle of judgment than the ‘typified’ form of the universal law formulation. That form of lawfulness will be a useful tool, or *heuristic*, but it will not be *constitutive* of judgment.

Reading Kant as tending towards the priority of the self, we can, finally, see how conscience supports an inference to freedom akin to the inference to transcendental freedom in the argument of the second *Critique*. The fundamental concept of conscience is the basic capacity to impute actions to ourselves, to *take responsibility* for our actions *and* for our dispositions. Through conscience I am responsible for the disposition to give priority to moral reasons in deliberation and action, a point which, taken to its limit, becomes the idea that I should take responsibility for my sensible nature itself. This responsibility implies an operation of spontaneity that is prior to the division into lawfulness and sensible inclination. Conscience as apperception is an activity that must be presupposed for the idea of duty to have any grip on the individual. My last and most radical Fichtean claim is that this Kantian conscience licenses an inference to a synthetic a priori act of self-constitution. This is what Fichte calls the absolutely self-active I, or intellectual intuition, the pure concept of activity that makes possible the construction of concepts of duty and right.⁷⁵ Only with such a concept can practical freedom, and the familiar, ordinary phenomena of conscience, be something more than dogmatic

⁷⁴ *Fichtes sämtliche Werke*, III, p. 10; *Foundations of Natural Right*, ed. Frederick Neuhouser and trans. Michael Baur (Cambridge: Cambridge University Press, 2000), p. 11.

⁷⁵ See Fichte’s polemical discussion of philosophers who take space and time as the only forms of intuition: ‘their “right” would have to turn out to be, let us say, square, while their “virtue” would perhaps have to be circular... But they do not really conceive of right or virtue in this manner. What then is it that underlies their construction of these concepts? If they observe correctly, they will discover that what underlies these concepts is acting as such, i.e., freedom’.

assumptions about human nature.⁷⁶ Though this argument does not establish transcendental freedom in exactly Kant's sense, it is enough to ground practical freedom and the autonomy of the normative realm.

I have argued that the potential gap between pure self-consciousness and law-governed self-imputation demarcates the central instability in Kant's conception of autonomy. My argument has examined Kant's difficulties in uniting the emphatically first-personal character of conscience and the objective constraint of the form of lawfulness. Of course Kant has strong reasons stemming from the metaphysics of transcendental freedom to take the form of lawfulness as constitutive of pure practical reason. Implicit in my argument has been the claim that moral philosophy should be guided by the dynamics of deliberation rather than by the metaphysics of the causality of the will. Within deliberation, the tension or instability becomes clear when one's paradigm cases are even moderately complex in the sense of involving more than one ground of obligation. In such cases, identifying with a moral judgment simply cannot be figured on conformity with a law or rule. There are, admittedly, serious problems looming for the priority-of-self view that I have been defending, most notably the possible 'subjectivism' of a first-person morality. But we cannot avoid those problems by asserting a separate set of objective faculties and objective constraints. Even if I do think of myself as doubled in conscience, my deliberation is a unitary phenomenon. Any theory of autonomy must come to terms with how the individual judges and acts as an integrated whole.

Fichtes sämtliche Werke, I, pp. 467–68; *Introductions to the Wissenschaftslehre and Other Writings*, ed. and trans. Daniel Breazeale (Indianapolis: Hackett, 1994), p. 51.

⁷⁶ This is to turn a point that Hill makes on Kant's behalf against Hill himself: 'even though Butler claims that a person following conscience is [] "a law to himself", one might argue that his or her ultimate guide is the given "constitution of human nature", whose normativity seems to be accepted as a given natural fact...arguably there is a deeper notion of integrity attributable to persons faithful to the Kantian conscience'. Hill, 'Four Conceptions of Conscience', p. 308.

Copyright of *Journal of Moral Philosophy* is the property of Brill Academic Publishers and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.