

Conscience

Philosophical theories of conscience might be categorized under three headings: *moral knowledge* theories, *motivation* theories, and *reflection* theories. People speak of their conscience “telling them to do” such-and-such. Such talk might imply that conscience is a source of moral knowledge about what to do. *Moral knowledge* theories of conscience try to explain how conscience affords us such knowledge. Some religious theories of conscience, for instance, interpret the voice of conscience as the voice of God within us. People also speak of “prickings” and “proddings” of conscience, or of their conscience urging them to do the right thing or bothering them if they have done (or are thinking of doing) the wrong thing. This suggests that conscience *motivates* us to do the right thing and to avoid the wrong thing. Conscience seems also to involve a certain way of *thinking reflectively* about what to do. This is usually a way of reflecting that gives first priority to moral considerations.

The three kinds of theory are not mutually exclusive. For example, Christian scholastic theories of conscience often distinguish *synderesis* (a notion derived from St. Jerome), which is a supposed source of moral knowledge, from *conscience*, which for some (e.g., St. Bonaventure) is an affective or volitional response to moral knowledge, while for others (e.g., St. Thomas Aquinas) it is the application of moral knowledge to action.¹ St. Bonaventure’s theory, therefore, combines knowledge and motivation.² An early modern example of a reflection theory is that of Joseph Butler, for whom conscience (also sometimes given the names “reason” and “the moral faculty”) consists in calm, rational reflection on what we ought to do.³ Reflection theories obviously don’t deny that both knowledge and motivation are needed for moral action, but they regard these as either presupposed by the reflection of conscience or else subsumed by it. An example of a pure motivation theorist would seem to be John Stuart Mill, for whom conscience consists in a painful feeling associated through our moral education with what we have been taught is the violation of duty (Mill, pp. 28–9). The

association, he thinks, tends with the intellectual progress of humanity to be diminished in motivational power by “the dissolving force of analysis,” so that for Mill, it would seem that the ultimate sanction of the utilitarian morality is (or, with further intellectual progress, will eventually be) not conscience at all but “the social feelings of mankind” (Mill, pp. 31–4).

In the terms I have just been using, Kant’s conception of conscience is a motivation theory set in the context of a reflection theory. Kant regards conscience as distinct both from our awareness of moral principles (through *practical reason*) and from the *faculty of moral judgment*, which are the sources of the moral knowledge that is to be implemented in the process of self-examination (and associated motivation) that is conscience (MS 6:438). In the *Metaphysics of Morals*, Kant treats conscience under two main headings: (1) as one of the moral feelings presupposed by our susceptibility to duty (MS 6:400–1) and (2) as a crucial aspect of a fundamental duty to ourselves, the duty of self-examination and self-knowledge as our own moral judge (MS 6:437–42). Kant’s principal theory of conscience is (2) the process of self-examination and self-judgment, but it will help us understand his moral psychology better if we begin with (1) conscience as a morally motivating feeling.

1. Conscience as Feeling

As we saw in Chapter 2, and again in the two preceding chapters, Kant is misread when he is seen as denying any role to feeling, emotion, or desire in those actions done from duty that have authentic moral worth. Moral action proceeds from desires produced in us by rational choice, grounded on principles of reason. In sensible creatures such as human beings, these purely rational desires, like inclinations, manifest themselves in the form of feelings – feelings resulting directly from the operation of reason on our sensibility. Susceptibility to these feelings is essential to our capacity to act rationally, and a being that was not susceptible to them could not be a responsible moral agent at all. In the *Groundwork*, the feeling of this kind that Kant emphasizes is *respect*, especially respect for the moral law (G 4:401). As we have already observed more than once, Kant distinguishes four distinct feelings that arise from pure reason: (a) moral feeling, (b) conscience, (c) love of human beings, and (d) respect (MS 6:399–403).

The last two feelings, (c) and (d), correspond to the two classes of our duties to others: duties of love and duties of respect (though the emphasis in that particular discussion of respect is on *self-respect*). (a) *Moral feeling* consists in a feeling of pleasure or displeasure (approval or disapproval) attached to actions, either performed or contemplated, and whether performed by another or by ourselves. (b) *Conscience* is a feeling of pleasure or displeasure associated with *myself*, in view of some action I am either contemplating or that I have already performed. In the former case, the feeling

is one that may motivate me either to perform the action or to refrain from it. In the latter case, it is a feeling either of self-contentment or of moral remorse. Because of its motivational force, Kant sometimes calls conscience as feeling an “instinct,” meaning that it is capable of impelling us to action and not merely of judging actions (VE 27:351). Conscience as a feeling, however, is the outcome of a specific process of moral *reflection*, and Kant’s proper theory of conscience consists in his account of this process.

2. The Inner Court

Kant’s theory of conscience, as a reflection theory, is characterized by Kant’s conception of conscience as an inner court or judicial proceeding:

Conscience is practical reason holding the human being’s duty before him for his acquittal or condemnation in every case that comes under a law. (MS 6:400)

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 judicial proceeding [*Gericht*] (*coram iudicio*), which, as a moral person giving effect to a law, is called a court [*Gerichtshof*] (*forum*). – Consciousness of an inner court in the human being (“before which his thoughts accuse or excuse one another”) is **conscience**. (MS 6:437–8)

The inner judicial proceeding of conscience may be aptly compared with an external court of law. Thus we find within us an accuser, who could not exist, however, if there were no law; though the latter is no part of the civil positive law, but resides in reason... In addition, there is also at the same time in the human being an advocate, namely self-love, who excuses him and makes many an objection to the accusation, whereupon the accuser seeks in turn to rebut the objections. Lastly we find in ourselves a judge, who either acquits or condemns us. (VE 27:354)

It is easy to take Kant’s talk about conscience as a court to be a mere metaphor, and in certain respects it obviously is. (There is no witness box, no judge’s gavel, no wooden-paneled room in a public building; and the judge – the moral agent himself or herself in a self-judging capacity – wears no robes.) But it is not as metaphorical as it might seem. The persons of the accuser, the defender, and the judge are literally accusing, defending, and judging, and though they are all the same natural person, they are literally distinct *moral persons* – as are the imposer and the subject of duties to oneself, and also the legislator and the subject of the moral law. To say that the court is “inner” means only that the relations between these persons are *self*relations.

For Kant, *innersense* is our intuition of our own conscious states rather than of objects distinct from us; *innermoral* worth is the worth we have in relation to our own self-given moral law rather than in comparison to other people. In the same way, the different persons involved in the inner court are all roles played by a single moral agent in a process of self-reflection.

In one way, however, it would seem that the image of a criminal court, taken literally as a forum for determining only guilt or innocence, seems inadequate to capture everything Kant needs conscience to do. Conscience, as pronouncing guilt or innocence, cannot be the only kind of moral self-reflection, however, because some acts can be morally meritorious, and the judgment as to whether they are is not a judgment of guilt or innocence. This is a modification of Kant's conception of conscience that seems to be required in Kantian ethics merely in order to make it self-consistent.

Kant's "legalisms." Whether or not the image of a court is a metaphor, the fact remains that Kant chooses this image quite deliberately, and this choice is worth some reflection. In every area of philosophy, not only in ethics but even in theoretical philosophy, Kant habitually uses metaphors, images, and analogies derived from laws and legal processes. In the title of the *Critique of Pure Reason*, the word "critique" is based on the Greek word for "judge," and in the preface to the first edition, Kant describes the "Critique of Pure Reason" itself metaphorically as a "court of justice" (*Gerichtshof*), before which "reason may secure its rightful claims while dismissing all its groundless pretensions, and this not by mere decrees but according to its own eternal and unchangeable laws" (KrV Axi–xii). The principle of morality too, as we have seen, is for Kant a "moral law," which we may regard as legislated within each of us by our own reason. Even the crucial idea of "judgment," which plays such a vital role in Kant's epistemology, involves reference to the image of a judicial forum.

Based on these metaphors, Kant's moral philosophy is sometimes criticized as "legalistic." As we saw in the previous chapter, however, Kant draws a sharp distinction between right, which is concerned with protecting external freedom and admits of coercive enforcement, and ethics, which is concerned with human perfection and happiness and involves only inner self-constraint. He is careful to observe the distinction between standards of right and standards of ethics in his moral philosophy, even if his interpreters too often ignore the distinction and run roughshod across it. Kant's use of legal and juridical metaphors is better seen as an expression of the values that would be displayed in an ideal judicial process: A judicial proceeding should be a public forum in which important matters are to be decided freely and fairly according to objective standards, with all sides being given the best opportunity to present their case. Kant's attraction to these metaphors depends on the open-mindedness and freedom of the process of inquiry, the objectivity and universality of reason's standards, and the importance, to any exercise of reason, of public communication about these objective

standards, operating under their authority. The same values also determine Kant's basic theory of conscience as an inner court of moral judgment. For Kant, the moral reflection of conscience must be rational, not merely a response to inchoate, prerational (or still less socially conditioned and inculcated) feelings. Conscience as feeling is to be the response of our sensibility to reason. Both the accuser and the defender within us must be seen as articulating their arguments on explicit grounds, and the verdict of the judge must equally be a reasoned one. The standards of argument are to be objective and universal, fair to both the prosecution and the defense, and the judge within us is to follow this law with integrity, allowing neither irrational self-hatred nor coddling favoritism to oneself to influence the decision. It is easy to deride Kant's use of judicial imagery, based on the pompous formalism of courts of law and the fact that in the real world they often perpetrate injustice rather than do justice. But once we understand the substantive values that underlie Kant's use of such images, it would be no joke to treat these values as lightly as our derision is in danger of doing.

Conscience and society. Kant is aware of theories of conscience that treat its voice as merely that of society, or of "art and education," but he distinguishes conscience in that sense (calling it *conscientia artificialis*) from self-judgment based on genuine rational principles (*conscientia naturalis*) (VE 27:355–6).

Insofar as what a person calls his "conscience" reflects only what society and upbringing have instilled in him, it can claim no rational or moral validity. Such a theory might understand conscience as a form of contingent, external motivation to morality (as Mill apparently does). In that case, however, the theory cannot also regard conscience as playing any role in moral *rationality* (as Kant, along with the entire tradition that includes Bonaventure, Aquinas, and Butler, thinks it should do). Because the court of conscience is an inner one, it is not literally a public forum but the moral law that all the inner parties recognize as one that has been legislated by the idea of the will of every rational being, and in that sense the rational standards used in the inner court are the same as would apply in a public forum.

In this respect, Kant's conception of conscience stands in striking contrast, for example, to the conception of conscience found in Martin Heidegger's in *Being and Time*. For Heidegger, the call of conscience is not articulate, or properly articulable, but consists in a discourse of silence. Conscience for Heidegger is Dasein calling itself back out of the public realm – which Heidegger dismisses in this context as merely the realm of *das Man* – to its own uncanny authenticity that recognizes no public standards.⁴ Heidegger explicitly considers, and rejects, Kant's image of conscience as a court, precisely on the ground that conscience does not put a "self" up for trial according to an explicit norm. That is why its discourse can involve no articulate utterances but consists only in "keeping silent" (*schweigen*).⁵ This is a deliberate rejection of the entire Enlightenment rationalist view of human

life (discussed in Chapter 1, §4), according to which the ultimate norms to which we are answerable are always to be seen in an ideal context of free human communication according to universal standards.

3. Conscience, Guilt, and Punishment

Conscience as a duty. We saw in the previous chapter that Kant regards conscience as our primary duty to ourselves in regard to our moral predisposition to personality. This duty to oneself, “the duty to oneself as one’s own innate judge,” Kant regards as fundamental to morality as a whole and to our observance of all ethical duties whatever class they may belong. “The human being has a general duty of so disposing himself that he may be capable of observing all moral duties . . . This, then, is the primary duty to oneself” (MS 6:348). It is in virtue of this duty that Kant regards duties to oneself as taking “first place” and as “the most important of all” (MS 6:341).

There is no duty to *have* a conscience, Kant argues, because unless we do, we are not moral beings at all and cannot be held responsible for our actions (MS 6:400). “Having a conscience,” in this sense, seems to mean having the fundamental *capacity* to carry out the kind of moral reflection conscience consists in. This capacity, for Kantian ethics – like the susceptibility to certain feelings and emotions, such as respect and love of human beings – is essential to human rationality itself. Our duty in regard to conscience consists in constraining ourselves to exercise these capacities and then attend to the verdict of our conscience. Of course many human beings fail to do this sometimes, or even all of the time. The duties involving conscience appear to be narrow or perfect duties. That is, it is not morally optional that we exercise the capacity for self-judgment whenever we face (or have made) a morally significant decision. Rather, in every morally significant choice, we are required to place ourselves before the inner judge and heed the verdict of the judge. To do this is not meritorious, but to fail to do it is blameable. Conscience has two distinct functions regarding our actions. It *warns* us (before we act) and it *pronounces a verdict* (of guilt or acquittal) over the actions we have already performed (MS 6:440).

Does conscience punish? The thought that conscience, or I myself as self-judge, sometimes renders a verdict of *guilty* might make us think that Kant would view us also as having the duty to punish ourselves for our misdeeds (as by depriving ourselves of the happiness of which we judge ourselves unworthy).⁶ Kant says that conscience may “judge us punishable” (or not, if it finds us not guilty), but he insists that our happiness or misery is left for the ruler of the world to decide (MS 6:439n, 440, 460). It would even be a fundamental misunderstanding of Kantian ethical theory, and of the role in it of the idea that morality is the condition of worthiness to be happy, to think that we are required or even entitled to punish ourselves. Kant holds that it can *never* be our duty to deprive ourselves of happiness (whether we

judge ourselves worthy of it or not) as long as no direct violation of duty is involved in the happiness itself or in the means of acquiring it. If we think we enjoy (or are hoping for) some happiness of which our conduct has not made us worthy, our only duty is to strive to improve our conduct so as to make ourselves less unworthy of that happiness.

Kant denies that it is even *possible* to punish oneself (MS 6:335). When he says this, I do not think he means to deny that people can deprive themselves of happiness or inflict suffering on themselves while thinking that they are undergoing punishment at their own hands. He does mean to deny, however, that such conduct could ever constitute *genuine* punishment, that it could count as an act rationally justified in the way legitimate acts of punishment are justified. Kant condemns the religious practice of penance, for example, as “slavish” and “hypocritical” (R 6:24n). It is slavish because it depicts us as trying to win the favor of the divine being by irrational acts of sycophancy. This is the only way Kant can understand such acts, because he thinks it is never our duty to deprive ourselves of any happiness unless the happiness itself or the means to it violate the moral law. It is also hypocritical because it is contrary to reason to deprive oneself of any happiness that does not directly involve immorality in its acquisition, and so human beings cannot honestly and wholeheartedly will to deprive themselves of happiness, even if they pretend to do so in their shameful attempts to humiliate themselves before God (whose goodness they also insult by supposing that this degrading behavior pleases him).

The only suffering we rightly undergo before the inner court of conscience consists in the painful feeling – a moral feeling, not an empirical one – that arises necessarily from the influence of reason on sensibility, attendant on the recognition that we have violated the moral law. But that pain is inseparable from the judgment itself, which is why conscience is counted among the feelings we can have no duty to have, because susceptibility to it is a presupposition of being morally accountable at all (MS 6:400–1). For Kant, however, *punishment* would have to be some *further* pain whose infliction is somehow due to us in consequence of the fact that we have done something wrong (MS 6:331). Or as Kant says: “*Moral remorse is the first outcome of the legally binding judicial verdict* [of conscience] . . . [But] even in *foro humano*, guilt is not assuaged by remorse, but by payment” (VE 27:353–4).

Kant’s view on this point, it seems to me, contains an important insight into what it means to have guilt feelings. To *feel guilty* is to *judge oneself punishable* – but it is not to *undergo punishment* or to inflict it on oneself. (Guilt feelings are painful, but it is an error – sometimes even a piece of culpable self-deception – to represent this pain as *punishment*.) On the contrary, our guilt feelings are sometimes assuaged when we are punished, for then we think justice has been done to us and we need no longer judge ourselves punishable. Kant’s view also affords us a way of drawing the distinction between guilt and shame. Moral shame, at any rate, is the feeling that we have failed

to live up to our humanity; it is what Kant means when he speaks of our feeling “self-contempt and inner abhorrence” (MS 6:426). This feeling might be prompted not only by the verdict of conscience that holds us guilty of some transgression of duty but also by awareness of a mere lack of moral merit in our character, a bad moral disposition, or even the absence of any good disposition. Guilt, by contrast, is the judgment that we have committed some actual deed that violates the moral law and renders us punishable.

4. The Duty of Self-Knowledge

The first command of duty regarding conscience, Kant says, is to “know (scrutinize, fathom) yourself” regarding your own maxims and the incentives on which you act (MS 6:441). This is a duty Kant regards as impossible to fulfill completely. As civilization (or modern bourgeois society) has made us, our developed reason and self-conceit have made us skilled in all forms of flattering self-deception, and in any case the truth about ourselves is often too abysmal for us to face. Even the striving after self-knowledge is attended with some serious dangers. One of them is “enthusiastic contempt” for oneself (or of the entire human species), leading either to fanatical self-hatred or to a misanthropy that violates our duty to promote the welfare of others. The antidote to it is keeping alive our awareness of the moral predisposition in us (the absence of which would signify not evil but simply a lack of moral personality altogether) (MS 6:441).

Kant’s target here is the morose self-scrutiny of certain religious self-examiners (such as Haller and Pascal) that leads sooner to madness than to truth (VA 7:133). This morbid attitude is closely allied in Kant’s mind with the pietistic religiosity in which Kant himself was raised, which “reduces [the moral agent] to a state of groaning passivity, where nothing great and good is undertaken but instead everything is expected merely from wishing for it” (R 6:184). The contrary danger – which actually bears a striking resemblance to its opposite – is the “egotistical self-esteem that takes mere wishes – wishes that, however ardent, always remain empty of deeds – for proof of a good heart” (MS 6:441). The self-knowledge Kant insists is a duty must avoid both these extremes. It is the sober resolve, as far as we are able, not to deceive ourselves about our deeds or about their sources within us and seeks a knowledge whose sole aim is constructive moral self-improvement.

Can conscience err? One arresting claim Kant makes about conscience is that an *erring conscience* is impossible. The question of whether conscience can err is often raised by moral philosophers in the tradition in which Kant is working, and there is great reluctance to admit that an erring conscience is possible. But there are also large differences over the question of what an erring conscience would have to consist in. In Kant, Fichte, Fries, and Hegel, for example, we find four different conceptions of what an erring conscience would have to be, and thus four different propositions asserted

as each philosopher denies that conscience can err.⁷ Because Kant regards conscience as distinct from moral judgment, he can (and does) hold that this judgment can err without holding that conscience errs. “An erring conscience,” he declares, is “an absurdity.”

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 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, then I would have made no practical judgment at all, and in that case there would be neither truth nor error. (MS 6:401)

According to Kant, it is quite possible for me to err in my objective moral judgments – thinking, for instance, that it is my duty to fight in a war, or to refuse to fight in it, when in fact my duty is just the opposite of what I think it is. This may also result in a conscience that condemns us based on false judgments (such as overly demanding standards – a notion some people may be surprised to find Kant employing at all). Kant calls this a “morbid conscience”: “But there is also a morbid conscience, where [the human being] seeks to impute evil in his actions, where there is no ground for it; but this is pointless. Conscience should not be a tyrant within us. We can always be cheerful in our actions, without offending it” (MS 6:356–7).

Because Kant does not identify conscience with moral judgment, he declines to infer from such cases that *conscience* can err. Conscience is the process of moral reflection that makes use of such moral judgments in delivering on myself a verdict of guilt or acquittal for some action I have done or am contemplating. The duty of conscience is therefore the duty to engage in this reflection. Its judgment proper is that I have applied the standards of moral judgment to myself (whether or not I have rendered the right substantive judgment in doing so). An errant judgment of guilt or acquittal would be an error of understanding, or of practical judgment, not of conscience. An *erring conscience* would have to be a mistaken judgment concerning whether I have held my actions up to the rational standards of moral judgment: It would have to be, so to speak, the mistaken judgment that the inner judicial process has taken place. It is this error that Kant apparently regards as impossible.

Conscience can also be defined as *the moral faculty of judgment, passing judgment on itself*... Conscience does not pass judgment on actions as cases that stand under the law, for this is what reason does so far as it is subjectively practical... 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 the human being to himself to witness for or against himself whether this has taken place or not. (R 6:186)

This seems clear enough, but in some ways it is hard to reconcile with what Kant plainly believes about people. Kant certainly realizes that many

people do not submit their actions to such a process of self-judgment, and he also often insists that people are extremely prone to self-flattering and self-exculpating illusions of all kinds. One such illusion would obviously be that I have submitted my act to the judgment of conscience when I have not. ("Hypocrisy" in the literal, etymological sense – that is, "deficiency in judgment" – might be regarded as the failure to pronounce conscientious judgment on one's actions when one ought to, while cherishing the illusion that one has properly judged them.) So it would seem quite inconsistent of Kant to deny that errors of this kind can occur. He knows they occur all the time:

There are tendencies in the souls of many to make no rigorous judgment of themselves – an urge to dispense with conscience. If this lack of conscientiousness is already, in fact, present, we never get that person to deal honestly with himself. We find in such people that they are averse to any close examination of their actions, and shy away from it, endeavoring, on the contrary, to discover subjective grounds on which to find a thing right or wrong. (R 6:616–17)

Kant is especially sharp in condemning the habit of thought that self-deceptively confuses the self-reproach-based imprudence with the verdict of conscience (R 6:24, VE 27:352–3). On the one hand, we perpetrate such confusions in the course of misinterpreting the moral law to ourselves so as to adapt it to our inclinations (G 4:405). On the other hand, we lend a certain air of dignity to our foolish (imprudent) acts by representing them as moral transgressions. Kant thinks that this is often a major ingredient in the hypocritical frame of mind of the self-torment of the sinner who inflicts religious penance on himself (R 6:24n). In both cases, we substitute a judgment of prudential reason for the verdict of conscience while persuading ourselves that we have made a conscientious examination of ourselves.

In view of all this, how can Kant consistently maintain that we cannot err in our judgment that I have submitted myself and my action to my practical reason in its role as judge? I suggest that what Kant might mean in denying an erring conscience is not that we cannot deceive ourselves in thinking that we have properly judged our actions when we have not but rather that if we do in fact genuinely submit ourselves to the judgment of conscience, then we cannot fail to be aware of doing so: "for [he says] 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" (MS 6:401). 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. In other words, where conscience has actually operated, we cannot be mistaken in thinking that it has. That is the only sense in which conscience cannot err.

The duty to pass judgment on oneself in conscience is for Kant our most fundamental duty. Without fulfilling it, we cannot honestly represent

ourselves either to others or to ourselves as having fulfilled any of our duties. Conscience would be unnecessary for an innocently good being, one that acts by a kind of inborn good nature or through some sort of training or upbringing, as a dog might be taught to do tricks. Such a being could also not rationally judge its actions good, and this means it could not be considered a moral being at all. It is an important claim of Kantian moral anthropology that the very social and historical processes that develop our capacity for moral reason also involve us in moral corruption. We never begin as moral agents with a clean slate, and our most fundamental moral action must always therefore be to struggle against ourselves (R 6:44–5, 72, 93–5). The most indispensable element in this struggle is the capacity for *conscience* – that is, honest self-judgment. We begin to see here why *truthfulness* – with others, but even more with oneself – is a fundamental *ethical* duty in Kantian ethics. (This is a theme to which we will return in Chapter 14, §4.)