

KANT ON CONSCIENCE

To be published in *Kantovski Sbornik*, in the
Russian translation by Leonard Kalinnikov

Allen Wood

allenw@stanford.edu

Philosophical theories of conscience may be categorized by bringing them 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 their conscience urging them to do the right thing, or bothering them if they have done (or are thinking of doing) the wrong thing. There are also “prickings” and “proddings” of conscience. This suggests that conscience motivates us to do the right thing and to avoid the wrong thing. Motivation theories attempt to account for this. Conscience seems also to involve a certain way of thinking reflectively about what to do – a way that gives first priority to moral considerations. A reflection theory of conscience is one that gives prominence to this aspect of conscience and tries to account for the kind of reflection it involves. Because moral epistemology, the psychology of moral motivation and the nature of moral reasoning are all fundamental topics in moral philosophy, these theories of conscience get at some of its central concerns.

These 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 ([19], I-II, I). St. Bonaventure’s theory, therefore, combines knowledge and motivation ([2]. Book II, Distinction 39). 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],

Sermons II and III). 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 under 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 ([18], pp. 28-29). 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" ([18], pp. 31-34).

In these terms, Kant's conception of conscience is a motivation theory set in the context of a reflection theory. Kant distinguishes conscience from moral principles and moral judgment, which are presupposed by it, but are identified with practical reason rather than with conscience. 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 ([13], pp.:400-401). 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 ([13] pp. 437-442). Kant's is definitely not a moral knowledge theory, since he regards conscience as distinct from the *faculty of moral judgment*, which is as the sole source of the moral knowledge that is to be implemented in the process of self-examination (and associated motivation) that is conscience ([13] p. 438). 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 predisposition to feeling which is an indispensable part of being a rational moral agent.

Conscience as a presupposition of morality on the side of feeling. Early in the *Groundwork*, Kant denies that authentic or genuine moral worth can belong to actions that are motivated by inclination (empirical desire) or by empirical feelings, such as sympathy ([10], pp. 397-399). These denials are often misunderstood in the most disastrous manner possible, when they are taken to deny that any moral value whatever could attach either to actions we want to do or to actions whose motivation involves feeling of any sort. This leads to the common misperception of Kant as a moral

philosopher who altogether hates and despises the sensitive or emotional side of our nature and thinks that morally good actions can be only those we have no desire at all to do.

Several distinct misunderstandings of Kant are involved in this picture, but the one that matters most for our present purposes is the ridiculously false idea that for Kant, action with genuine moral worth must be unaccompanied by either feeling or desire. On the contrary, Kant's psychology of action involves the thesis that all action involves the representation of an end to be produced, and a desire for that end, and the conception of desire for an end as the representation of it accompanied by a feeling of pleasure (if the feeling is displeasure, then it is a case not of desire but of aversion). The key point to understand is that for Kant, not all desire is inclination. 'Inclination' refers only to habitual *empirical* desire. Fundamental to Kant's psychology of action (also to Aristotle's, with which Kant's is often mistakenly contrasted at this point) is the idea that not all desire arises passively through our receptiveness to empirical impulses, but some desires can result solely from the activity of pure reason, the free and self-directing side of our nature. (To assume that all desires are inclinations, and that all feelings are empirical in origin, is to deny Kant's entire moral psychology from the start. To interpret Kant while making such assumptions is to guarantee that the position one ascribes to him will be utterly nonsensical.) In sensible creatures such as human beings, purely rational desires, like empirical desires, also 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 who 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 ([10], p. 401). Perhaps respect seems to us like such an austere feeling – in contrast to sentiments such as love and sympathy -- that we tend to overlook the fact that it is a feeling at all; that error on our part may account for the fact that readers of the *Groundwork* can be aware of the crucial role Kant assigns to this feeling, and yet at the same time draw the conclusion from this text that moral action for Kant involves no feeling at all. In the *Metaphysics of Morals*, however, Kant distinguishes four distinct kinds of feelings that arise from pure reason, all of which are

essential to being a rational moral agent at all. They are (a) moral feeling (b) conscience (c) love of human beings and (d) respect ([13], pp. 399-403).

The last two feelings (c) and (d) would seem to correspond to the two classes of our duties to others: duties of love and duties of respect. That Kant regards a kind of love as necessary for moral agency may come as a surprise to some, especially if their impression of Kant's views about love have been shaped by the mistaken idea that all love for Kant must be either "pathological love" (which is a feeling and cannot be a duty) or else "practical love" (which is a duty and cannot be a feeling) (see [10], p. 399, cf. [11], pp. 76-77). Love of human beings, however, is a *feeling* (though not an empirical one, arising from sensibility, but a rational one – an immediate effect of reason on our sensibility) and it cannot be a duty, because our susceptibility to it is a presupposition of having any duties at all ([13], pp. 400-401).

Moral feeling (a) 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. Conscience (b) is a feeling of pleasure or displeasure associated with *myself*, in view of some action I am either contemplating or which I have already performed. In the former case, the feeling is one which may motivate me either to perform the action or to refrain from it; latter case, it is a feeling either of self-contentment or of moral remorse. All four feelings belong essentially to our emotional equipment, so to speak, as rational and moral beings. We can have no duty to acquire the susceptibility to any of them, because without this, we would not be persons (morally responsible agents) at all, and would therefore not be subject to any duties. Our duties regarding them can consist only in duties to be responsive to them, and to cultivate them rather than permitting them to be dulled or overwhelmed by contrary impulses or inclinations.

The feeling (or feelings) of conscience, constitute the *motivational* aspect of conscience in Kant's theory. Because of its motivational force, Kant sometimes calls conscience an "instinct", meaning that is capable of directing us to action, not merely of judging actions ([16], p. 351). The feelings and motivations that pertain to conscience, however, are the outcome of a specific process of moral *reflection*, and it is in his account of this process that Kant's proper theory of conscience consists. So let us now turn to it.

Conscience as an inner moral court of judgment.

“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” ([13], p. 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** ([13], pp. 437-438).

“The inner judicial proceeding of conscience mayh 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” ([16], p. 354).

It is natural to take Kant’s talk about conscience as a court to be a metaphor, and in certain respects it obviously is metaphorical. (There is no witness box or gavel, no wooden-paneled room in a public building, and the judge, who is the moral agent himself or herself in a self-judging capacity, wears no robes.) But it is important to realize that in at least one important respect, it is not as metaphorical as we might think. The persons of the accuser, the defender and the judge are, I think, quite literally distinct moral persons,

whose interaction, in Kant's view, typically follows a pattern very similar to that of the prosecutor, the defense attorney and the judge in an external court.

To say that the court is "inner" means that the different persons involved in it (the "accuser", "defender" and "judge") are all found within a single moral agent. The idea that a human being can contain distinct "persons" in this sense is basic to Kant's conception of moral agency and to moral legislation as self-given (autonomy). To be a morally self-directing rational being for Kant entails that there be within you both a legislator (an author of obligation, which gives the law) and a subject of the legislation (which is obligated by it). Thus it is a distortion of Kant's theory of autonomy to think that it is we finite, fallible human beings who give the law to ourselves, whether this misinterpretation takes the form of the Romantic notion that the moral law proceeds from the creative will of each human individual (and is therefore irretrievably individual in its content) (Schlegel, [1], pp. 132, 155, Schleiermacher [1], pp. 174-175), or the more recent form in which the supposedly Kantian doctrine is that it is we who "confer value" on objects by the choices we make ([17]). For Kant, the *author* of moral laws is "the idea of every rational will as giving universal law" ([10], pp. 431, 432), that is, the pure rational concept of each of us, which cannot be exhibited empirically. The subject of the moral law is the finite, empirical will of each of us, which is bound by this law. It is only through this dual personality, so to speak, that Kant thinks we can make sense of the notion of moral obligation, especially of obligations to oneself – since unless the author and subject of the obligation are different persons, the subject of the obligation could always release himself from it, and so the obligation would always be null and void ([13], pp. 417-418). The notion that natural persons (human individuals) do not correspond one-to-one to moral persons is not a novel one with Kant ([9], pp. 111-115).

The specifically Kantian thesis that the same natural person can combine more than one moral person is essential to his conception of autonomy, which will either be rejected from the start or badly misunderstood if this thesis is not accepted. Kant's theory of conscience is yet another application of the same thesis, this time as a theory not of moral legislation but of the kind of moral reflection in which conscience consists. Although some acts can be meritorious, in Kant's view conscience is not a kind of reflection in which the issue of merit arises. (Conscience is therefore not the only kind of moral

reflection, though it is in Kant's view the most fundamental and morally important kind.) Conscience is always a reflection on one's actions in which the issue, as in a criminal court, is guilt or innocence. This too is meant literally, not metaphorically, even though in the case of conscience the law is not an external or coercive one, but an ethical law, whose obedience must always be a matter of inner self-constraint.

Whether the image of a court is a metaphor or rather in certain respects a literal similarity, 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 and analogies derived from laws and legal processes. Sometimes he may even seem to be obsessed with them. 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" ([12], Axi-xii). The principle of morality is for Kant a "moral law", legislated within each of us by our own reason.

Based on these metaphors, Kant's moral philosophy is sometimes criticized as "legalistic," but of course Kant also draws a sharp distinction between the external legislation pertaining to right, which admits of coercive enforcement, and the ethical legislation of morality, which involves only inner self-constraint. A sympathetic interpretation of Kant's use of legal and juridical metaphors will highlight the fact that a judicial proceeding is 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 thus 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.

These values also determine Kant's basic theory of conscience as a metaphorical inner court of moral judgment. The moral reflection of conscience for Kant must be rational, not merely a response to inchoate, pre-rational (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, the law of reason, which are 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.

Kant is aware of theories of conscience that treat its voice as merely that of society, or “art and education,” but he distinguishes conscience in that sense (calling it *conscientia artificialis*) from self-judgment based on genuine rational principles (*conscientia naturalis*) ([16], pp. 355-356). Insofar as what a person calls his ‘conscience’ reflects only what society and upbringing have instilled in him, it has no rational validity. To hold that human beings never have any conscience except this artificial one would be, in effect, to deny that they are moral agents at all and it would follow that none of them could be held morally responsible for anything. If society’s customs involve something like holding people responsible, then they would be a mere sham, deserving of no rational respect whatever. In effect, then, to hold that all conscience is artificial in this sense would be for Kant to embrace an extreme skeptical position about all morality (and not merely to advance a theory about conscience).

Since the court of conscience is an inner one, it is not literally a public forum, but the moral law that all parties recognize is 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 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 as the realm of *das Man*) to its own uncanny authenticity, which recognizes no public standards ([7], p. 323). 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, and therefore its discourse can involve no

articulate utterances ([7], pp. 316, 318, 339). (Heidegger also denies that his characterization of conscience as a “call” is a “picture,” such as he takes Kant’s characterization of conscience as an inner court to be ([7], p. 316). But I have already suggested certain important respects in which the account of conscience as an inner court is not metaphorical – or, therefore, “pictorial” either.) Heidegger’s account of conscience displays in these ways a clear affinity with the Romantic conception of moral autonomy, for which the autonomous self is radically individualized, cut off from the public realm of rational discourse, and becomes a matter either of feeling or of some other kind of inarticulable awareness, such as that of the Heideggerian “uncanny.”

Conscience as a duty to oneself. The basic division among ethical duties for Kant is between duties to oneself and duties to others. This division is best understood in terms of the Formula of Humanity as End In Itself. A duty d is a duty to S if S is a rational being and d is grounded in the requirement to treat S as an end in itself. A duty is to oneself when it is one’s own dignity and worth as an end in itself that grounds the demand. Kant does not understand duties to oneself as duties to promote one’s own welfare, but rather as duties to act in such a way as to be worthy of one’s humanity. “Self-regarding duties do not depend on the relation of actions to the ends of happiness” ([16], p.:343). “All such duties are founded on a certain love of honor consisting in the fact that a human being values himself, and in his own eyes is not unworthy that his actions should be in keeping with humanity” ([16], p. 347). We have seen already how Kant deals with the potential problem that a duty to myself might be understood as placing me in a position to release myself from the duty whenever I like (thus rendering the duty null and void). Kant’s response to this problem is as radical as it is consistent with his basic conception of moral autonomy. Kant holds that the person who imposes the duty (the moral legislator in me) is not the same *person* as the subject of the obligation. We have seen that this conception of more than one person combined in the same human being is for Kant basic to the entire idea of a self-governing moral agent. The Kantian theories of autonomy and conscience, as well as the conception of a duty to oneself, is grounded on it.

Kant divides duties to oneself into duties to oneself as an animal being and as a moral being. Under the former heading, he places duties forbidding suicide, gluttony and

drunkenness, and sexual self-defilement ([13], pp. 421-428); under the latter heading he places duties forbidding lying, miserly avarice, and servility ([13], pp. 429-437). But there is one duty to oneself that Kant regards as fundamental to morality as a whole, and to our observance of all ethical duties -- to whatever class they may belong. This is the set of duties to oneself that one has as one's own innate judge. As we have seen, this is the duty associated with *conscience* ([13], p. 437-442). "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" ([16], p.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" ([16], p. 341).

It cannot be our duty to *have* a conscience, Kant argues, since unless we do, we are not moral beings at all and cannot be held responsible for our actions ([16], p.400). "Having" a conscience, in this sense, seems to mean having the capacity to carry out the kind of moral reflection conscience consists in, and to fulfill the roles of accuser, defender and judge of oneself. Our duty consists in constraining ourselves to exercise these capacities and then attend to the verdict of our conscience. This many human beings fail to do 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 when we exercise them, and there is nothing meritorious about exercising them. Rather, we are required with respect to all our actions to place ourselves before the inner judge, and heed the verdict of this judge; to fail to do so is always blamable. 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 ([13], p. 440).

Conscience, guilt and punishment. The thought that conscience sometimes renders a verdict of guilty might make us think that he 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). Kant does say that conscience may "judge us punishable" (or not), but he insists that our happiness or misery is left for the ruler of the world to decide ([13], pp. 439n, 440).

There is an insight here too, I think, about what it is to have guilt feelings. To feel guilty is to judge oneself punishable – but it is not to undergo punishment or inflict it on oneself. This is why our guilt feelings are assuaged when we are punished, since then we

think justice has been done to us and we no longer regard ourselves as punishable. This affords us also a Kantian 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” ([10], p. 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.

For Kant, however, we never punish ourselves, or others either, for *ethical* misconduct. Thomas Hill, ([8], Chapter 9), provides us with a very good discussion of Kant’s conception of conscience, but one point where he seems to me to go seriously wrong is in his claim that conscience involves the actual infliction of self-punishment ([8], p. 303). Kant denies that it is possible to punish oneself ([13], p. 335). He sharply condemns the religious idea of penance, for example, as “slavish” and “hypocritical” ([14], p. 24n). It is slavish because it depicts us as trying to win the favor of the divine being by morally indifferent acts of sycophancy. This is the only way Kant can understand it, since 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 thinking that this degrading behavior would please him).

The only “punishment” we suffer before the inner court of conscience is 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 ([13], pp. 400-401). For Kant, however, *punishment* would have to be some *further* pain whose infliction is some how

due to us in consequence of the fact that we have done something wrong ([13], p. 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” ([16], pp. 353-354).

It would 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 is *never* 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 acquiring it. Our only duty in this regard is to strive to improve our conduct, so as to make ourselves worthy of the happiness we both enjoy and hope for. Still less is it ever our task to decide when others are deserving of happiness. The happiness of others should always be among our ends. This forbids us from ever placing the unhappiness of another (for any reason) among our ends, as we would have to do if we set out to deprive them of some happiness of which we arrogantly deemed them unworthy. Once again, when we regard human beings as under *moral* laws (as distinct from external juridical laws that are punishable by the state), Kant holds that “God alone” is entitled to inflict punishment ([13], p. 460).

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 ([13] p. 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 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 not signify evil but simply a lack of moral personality altogether) ([13], p. 441).

Kant’s target here is the morose self-scrutiny of certain religious self-examiners (such as Haller and Pascal) which leads sooner to madness than to truth ([15], p. 133). This morbid attitude is closely allied in Kant’s mind to 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” ([14], p. 184). The contrary danger – which actually bears a strong 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” ([13], 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 striking claim that Kant makes is that an erring conscience is impossible. The question 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 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 (Fichte, [4], p. 174, Fries [5], pp. 214-215, Hegel [6], §§ 138-139).

Since 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” ([13], p. 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, 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). 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” ([16], pp. 356-357).

Since Kant does not identify conscience with moral judgment, he declines to infer from such cases that *conscience* can err. For Kant, conscience is rather 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 a kind of second-order reflection, judging that one has applied moral judgments properly to oneself.

“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” ([14], p. 186).

For conscience to err, therefore, would be for me to be hold mistakenly that I have submitted myself and my action to this process when in fact I have not. It is this error that Kant apparently regards as impossible. Kant, however, realizes that many people do not submit their actions to such a process of self-judgment, and 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 – meaning ‘deficiency in judgment’ would seem to consist in failing to pronounce conscientious judgment on one’s actions when one ought to, while cherishing the illusion that one has done so.) “There are tendencies,” Kant says, “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” ([16], pp. 616-617).

Kant is especially sharp in rejecting the confusion of self-reproach based on our imprudence with verdicts of conscience ([16], pp. 352-353). 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 ([10], p. 405). On the other hand, we lend a certain air of dignity 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 ([14], p. 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 suspect that what Kant means in denying an erring conscience is not that we cannot deceive ourselves about this, but rather that if we do in fact genuinely submit ourselves to the judgment of conscience, we cannot fail to be aware of doing so: “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” ([13], p. 401). In other words, where conscience is actually in operation, we cannot be mistaken in thinking that it is. And in that sense, conscience cannot err. But this would not preclude our mistakenly (or self-deceptively) thinking that conscience is operating when it is not.

The duty to pass judgment on oneself in conscience is for Kant the most fundamental duty in the sense that without fulfilling it, we cannot honestly represent ourselves either to other or to ourselves as having fulfilled any of our duties. We might of course still be a kind of innocently good moral agent who fulfills our duties instinctively or unreflectively by a kind of inborn good nature, or through some sort of training or upbringing that leads us to do the right by a kind of blind habit, as a dog might be taught to do tricks. It is an important part of Kant’s moral anthropology, however, that human beings are not spontaneously good in this way – that the very same social and historical processes that develop their capacity for moral reason also involve them in moral corruption, so that they never begin as moral agents with a clean slate, but as soon as they

become aware of themselves as moral beings, they must also be aware of themselves as morally corrupted, and their fundamental moral action must always therefore be to struggle against themselves ([14], pp. 44-45, 72, 93-95). For such beings as we are, therefore, conscience (self-knowledge and self-judgment) is a morally indispensable activity, lying at the very foundation of the moral life.

Bibliography

- [1] Beiser, Frederick (ed.), *The Early Political Writings of the German Romantics*. Cambridge: Cambridge University Press, 1996.
- [2] Bonaventure, *Commentary on the Sentences, Works*. St. Bonaventure, NY: Franciscan Institute, 1979). Cited by Book and Distinction.
- [3] Butler, Joseph, *Five Sermons*, ed. Stephen Darwall. Indianapolis: Hackett, 1983. Cited by sermon number.
- [4] Fichte, J. G., *System der Sittlichkeit, Sämmtliche Werke*, ed. I. H. Fichte. Berlin: deGruyter, 1970. Bd. 4.
- [5] Fries, J. F., *Handbuch der Praktische Philosophie, 1. Theil: Ethik*. Heidelberg: Mohr und Winter, 1818,
- [6] Hegel, G.W.F. *Philosophie des Rechts, Werke*. Frankfurt: Suhrkamp, 1970. Bd. VII. Cited by §.
- [7] Heidegger, *Being and Time*, translated by Robinson and Macquarrie. New York: Harper, 1962.
- [8] Hill, Thomas, *Human Welfare and Moral Worth*. Oxford: Clarendon Press, 2002.
- [9] Hobbes, Thomas, *Leviathan*. Cambridge: Cambridge University Press, 1992.
- [10] Kant, Immanuel, *Grundlegung zur Metaphysik der Sitten, Immanuel Kants Schriften*. Ausgabe der königlich preussischen Akademie der Wissenschaften. Berlin: W. de Gruyter, 1902-. Bd IV.
- [11] Kant, *Kritik der praktischen Vernunft, Kants Schriften*. Bd. V.
- [12] Kant, *Kritik der reinen Vernunft, Kants Schriften*. Bd. III-IV. Cited by A/B pagination in the original editions.

- [13] Kant, *Metaphysik der Sitten*, *Kants Schriften*. Bd VI.
- [14] Kant, *Religion innerhalb der Grenzen der bloßen Vernunft*, *Kants Schriften*. Bd. VI.
- [15] Kant, *Anthropologie in pragmatischer Hinsicht*, *Kants Schriften*, Bd. VII.
- [16] Kant, *Vorlesungen über Ethik*, *Kants Schriften* Bd. XXVII. Cambridge: Cambridge University Press, 1996.
- [17] Korsgaard, Christine. *The Sources of Normativity*. Cambridge: Cambridge University Press, 1996.
- [18] Mill, John Stuart, *Utilitarianism*, ed. George Sher. 2nd edition. Indianapolis: Hackett, 2001.
- [19] Thomas Aquinas, *Summa Theologia*. New York: McGraw-Hill, 1964-. Cited by part and question.

About the Author. Allen Wood is Ward W. and Priscilla B. Woods Professor of Philosophy at Stanford University. He has also taught at Cornell University (1968-1996) and Yale University (1996-2000) and has held visiting appointments at the University of Michigan (1973), the University of California at San Diego (1986) and Oxford University (2005). He is author of eight books, co-author of one more, and editor or co-editor of eight others. He is general editor, along with Paul Guyer, of the Cambridge Edition of Kant's Works in English Translation.