

Generating Correct Moral Judgments: Extending Kant's Categorical Imperative Procedure Beyond Rawls and Wood

Ziyu Zhu

School of Philosophy, Beijing Normal University, Beijing, 100091, China

Abstract: Kant's account of moral judgment reveals that the Categorical Imperative functions not as a corrective rule but as the law of judgment itself: correct moral judgments arise directly through testing maxims for moral worth. Allen Wood's realist reading misinterprets this relation by subordinating moral character to intellectual capacity and treating the imperative as a means of correction. John Rawls's constructivist interpretation captures Kant's procedural intent more faithfully, viewing moral judgment as the outcome of a four-step Categorical Imperative Procedure that tests whether a maxim can become a universal law. Yet a procedure grounded solely in the Formula of Universal Law cannot examine the moral motive or the autonomy of the will. Integrating Kant's Formulas of Humanity and Autonomy yields a six-step model that supplements universalizability and willability with two further criteria: the end-test, ensuring that humanity is treated as an end in itself, and the legislation-test, confirming that the will legislates universally and autonomously. This augmented structure preserves Rawls's constructivist framework while meeting Kant's demand that moral worth derive from duty, respect for humanity, and the self-legislation of reason, thereby providing a more coherent and textually grounded account of how correct moral judgments are generated within a Kantian framework.

Keywords: Kantian Ethics; Categorical Imperative; Moral Judgment; Constructivism; Autonomy.

1. Introduction

In recent years, debates between moral realism and constructivism have centered on how Kant conceives the generation of "correct moral judgments." Clarifying this issue not only illuminates the divergent interpretations of Kant's moral philosophy, but also deepens our understanding of its principle and practical import, preventing a misreading of the Kantian moral law. Allen Wood, representing realism, holds that moral judgments arise from ordinary moral reason but may deviate under temptation, thus requiring correction through moral principles. By contrast, the constructivist John Rawls argues that rational agents can construct correct moral judgments directly through a properly formulated Categorical Imperative procedure. Their dispute thus turns on how such judgments are to be produced.

This paper argues that Rawls's interpretation better reflects Kant's text, though it remains incomplete: a procedure based solely on the Formula of Universal Law still permits counterexamples. Drawing on Kant's own writings, I propose to revise and extend the Categorical Imperative procedure—without altering its structure—to address these deficiencies. The paper proceeds as follows. First, I reconstruct and critique Wood's account. Second, I examine Rawls's interpretation and show how the Categorical Imperative yields sound moral judgments. Finally, I analyze the sources of certain counterexamples and propose an augmented version of the procedure that resolves the framework's inherent limits.

2. Wood: The Categorical Imperative as a Correction of Moral Judgment

According to Allen Wood, a correct moral judgment arises not directly from human rationality but from the correction of

judgments produced by ordinary moral rational knowledge through the moral principle expressed in the Categorical Imperative. As he writes, "the task of FUL/FLN (the Formula of Universal Law and Formula of Law of Nature) is correcting our judgment when we are faced with that specific kind of moral difficulty." [1] Wood's interpretation rests on two central claims: (1) moral capacity operates only after intellectual capacity, and therefore moral principles function merely as correctives; and (2) when an agent experiences moral tension, moral principles serve as a means to escape from obligation.

Regarding the first claim, Wood distinguishes two capacities within rational beings—intellectual and moral. The former, which he calls "ordinary moral rational knowledge," is the intellectual expertise possessed by all rational agents. As Kant writes, "I need no well-informed shrewdness to know what I have to do in order to make my volition morally good." [2] Ordinary moral reason thus discerns right from wrong but tends to calculate outcomes and choose actions that appear most advantageous. The latter capacity, "moral character," represents the operation of pure practical reason, which constrains the will through moral law and compels us to act "for the sake of duty." Moral character, therefore, corrects the self-interested tendencies of ordinary rationality, functioning as a "moral compass" that steers agents back toward duty when they are tempted by inclination or self-interest.

For Wood, this correction occurs through the Categorical Imperative, which regulates the will and ensures that judgments gain moral worth. Its basic formulation commands: "Act only in accordance with that maxim through which you can at the same time will that it become a universal law." [3] Wood presupposes a temporal sequence between the formation of moral judgments and the moral principle's corrective intervention. As he puts it, "The problem FUL addresses is one that can arise only after reasoning – or

“common rational moral cognition” – has had its say.” [4] Accordingly, pure practical reason becomes effective only after ordinary moral reason has already produced a moral judgment. Thus, for Wood, the Categorical Imperative does not itself generate correct moral judgments; rather, it corrects the errors of ordinary reason. His reading is illustrated by Kant’s example of suicide. A person wearied by life may, guided by self-love, conclude that ending one’s life is the best way to escape suffering. Yet, when tested by the Categorical Imperative, the maxim “out of self-love, I take my own life” fails, since it “cannot become a universal moral law.” Hence, moral character—acting through the imperative—corrects the mistaken reasoning of ordinary moral knowledge, transforming a false judgment into a correct one.

Wood’s second claim concerns what he terms “moral tension.” He argues that in Kant’s examples, the agent’s perplexity arises not from abstract maxims but from a conflict between a definite duty and a temptation to act otherwise. Thus, the moral principle’s role is not to test a maxim but to correct the will that seeks to evade duty. The maxim merely reflects the agent’s inner conflict. On one side, ordinary moral reasoning tempts the agent to rationalize wrongdoing—for instance, justifying a “false promise” as a necessary means of relief. On the other, the Categorical Imperative exposes this rationalization by revealing its contradiction: “I thus transform this claim of self-love into a universal law and set up the question thus: How would it stand if my maxim became a universal law? Yet I see right away that it could never be valid as a universal law of nature and still agree with itself, but rather it would necessarily contradict itself.” [5] At this point, the imperative functions as a “law of judgment,” refuting the erroneous moral reasoning that seeks to justify wrongdoing. As Wood concludes, “moral agents don’t need philosophy to tell them what to do. It can offer moral agents only a canon of judgment, helping them to find their moral compass, so that they don’t self-deceptively misapply common rational moral cognition in a particular case.” [6]

Despite its coherence, Wood’s interpretation conflicts with Kant’s own text on two counts. First, his claim that moral character follows upon intellectual capacity is inconsistent with Kant’s view of moral law as an a priori fact of reason. For Kant, “the moral law is given, as it were, as a fact of pure reason of which we are a priori conscious and which is apodictically certain...the objective reality of the moral law cannot be proved by any deduction, by any efforts of theoretical reason, speculative or empirically supported...and it is nevertheless firmly established of itself.” [7] Moral principles are not empirically acquired correctives but intrinsic to rational agency. Therefore, moral character does not emerge after intellectual reasoning—it already operates within it. Thus, pure practical reason is not posterior to empirical practical reason; rather, it grounds it from the outset. The moral law lies latent within rationality itself, surfacing when inclination or interest produces conflict. Wood’s separation of moral and intellectual capacities thereby undermines the unity of reason and misrepresents Kant’s conception of moral law as a “fact of reason.”

Second, a return to Kant’s text reveals that the Categorical Imperative is not a corrective instrument but the very principle of moral judgment. Kant writes: “But one does better in moral judging always to proceed in accordance with the strict method and take as ground the universal formula of the categorical imperative” [8] Accordingly, “One must be able to will that a maxim of our action should become a

universal law: this is the canon of the moral judgment of this action in general.” [9] The imperative thus is the law of judgment that determines moral worth. There is, therefore, only one Categorical Imperative, namely: act only according to that maxim whereby you can at the same time will that it should become a universal law.

From this it follows that a correct moral judgment arises precisely when the maxim of an action can become, and is willed as, a universal law. The Categorical Imperative is not a posterior regulation meant to correct the will’s errors; it is the principle through which the will itself produces correct moral judgments. The maxim gains moral worth only when it satisfies the imperative’s universalizability condition. Therefore, Wood’s account misreads Kant by treating the imperative as an external corrective to ordinary reason, rather than as the internal rule of moral cognition itself.

In conclusion, Kant’s view implies that correct moral judgments are not derived from the correction of error but are generated directly by the Categorical Imperative. This reading accords more closely with John Rawls’s constructivist interpretation, which treats the Categorical Imperative as a procedural framework through which rational agents construct correct moral judgments. By situating the law within the very structure of reason rather than as a subsequent constraint, Kant anticipates the constructivist insight that moral objectivity arises from the rational procedure itself.

3. Rawls: The Categorical Imperative as the Generation of Moral Judgment

Rawls holds that correct moral judgments are generated through a four-step Categorical Imperative Procedure (CI Procedure) that tests the maxim of an action. If the maxim passes this test, the corresponding action is moral. He writes, “a correct judgment is one that conforms to all the relevant criteria of reasonableness and rationality the total force of which is expressed by the way they are combined into the CI-procedure.” [10]

For Rawls, the Categorical Imperative provides the standard of moral worth. As Kant states, “The categorical imperative is thus only a single one, and specifically this: Act only in accordance with that maxim through which you can at the same time will that it become a universal law.” [11] The moral law thus demands that one’s maxim be capable of universalization and that one will to live under such a universal order. Only an action guided by such a maxim can be law-like and possess moral worth. In other words, a correct moral judgment arises when the will acts according to a maxim that meets this law-likeness condition, tested through the CI Procedure.

How can such testing occur, given that the imperative is purely formal while maxims are empirical? Rawls answers that “I assume that categorical imperative is applied to the normal conditions of human life by what I shall call the CI-procedure.” [12] The procedure thus operationalizes the categorical imperative, providing a structure for determining whether a maxim could become a universal law. As he explains, it must have sufficient structure to specify the requirements of moral thought, so as to show that many maxims are or are not suitable to be universal laws. In this sense, the CI Procedure functions as a testing framework uniting pure and empirical practical reason: a maxim that passes is moral; one that fails is not.

Rawls insists that judgments produced through this procedure are correct because the procedure itself embodies “all the relevant criteria of practical reasonableness and rationality”. [13] Since it is derived entirely from the categorical imperative, any judgment emerging from its proper use already conforms to reason’s formal demands. The CI Procedure, therefore, is not merely a tool of evaluation but the very mechanism through which reason generates moral judgments.

Rawls formulates this mechanism in four steps:

(1). I am to do X in circumstances C in order to bring about Y unless Z. (Here X is an action and Y is an end, a state of affairs.)

(2). Everyone is to do X in circumstances C in order to bring about Y unless Z.

(3). Everyone always does X in circumstances C in order to bring about Y, as if by a law of nature (as if such a law was implanted in us by natural instinct).

(4). We are to adjoin the as-if law of nature at step (3) to the existing laws of nature (as these are understood by us) and then think through as best we can what the order of nature would be once the effects of the newly adjoined law of nature have had sufficient time to work themselves out. [14]

In Step 1, the agent formulates a maxim: action X for end Y in condition C. Step 2 universalizes it. Step 3 imagines this universalization as a law of nature. Step 4 integrates the hypothetical law with the existing order of nature to see whether any contradiction results. If the resulting world contains a contradiction—logical or practical—the maxim fails.

(1): I am to make a deceitful promise in circumstances C (that is, when I am in embarrassing straits and need money, even though I know that I cannot repay the debt, and have no intention of doing so) in order to further my own personal advantage.

(2): Everyone is to make a deceitful promise in circumstances C, etc., as above.

(3): Everyone makes (or tries to make) a deceitful promise in circumstances C, etc. (as if by a law of nature).

(4): Adjoin the law of nature at step (3) to other laws of nature (as known by us) and figure out the equilibrium state that would result. This adjusted social world is one in which no one can make a deceitful promise in circumstances C, as much as they would like to do so. [15]

The adjusted world that results is one in which, though many would wish to act thus, no one could in fact make a deceitful promise, since promises would lose all credibility. As Kant remarks, “No one would believe the promise of another.” In such a world, the very practice of promising collapses. Hence the maxim “to make a false promise for self-interest” fails the CI Procedure, for the law of nature at step (4) negates the universalized maxim at step (2): if everyone lies, no one can successfully promise. The resulting order is self-defeating and cannot be willed by rational agents.

In summary, Rawls derives from the categorical imperative a four-step mechanism that determines whether a maxim can be universalized. To act morally is to act according to a maxim that passes this test. Yet Rawls’s interpretation remains incomplete. Because it rests solely on the Formula of Universal Law, it cannot account for cases in which a maxim formally passes the procedure but still lacks moral worth. As Kant notes, “Finite beings experience the moral law as a constraint,” and true morality requires that the will act from duty rather than merely in accordance with it. Thus, while

Rawls captures the procedural logic of Kant’s ethics, he understates the role of motive and respect for law in constituting moral worth.

Rawls’s constructivist account nonetheless clarifies the rational structure of moral judgment. The CI Procedure translates Kant’s formal principle into a stepwise method linking pure reason and empirical practice. It shows how reason tests maxims through universality and thereby generates correct moral judgments. Yet because a purely procedural test cannot ensure that the agent acts from the motive of duty, the procedure alone is insufficient to capture the full sense of moral worthiness in Kant’s ethics. A complete account must therefore supplement Rawls’s formal construction with Kant’s notion of the moral law as both a principle of reason and a “fact of reason” that moves the will.

4. Extending the CI Procedure

In Rawls’s view, once a maxim has passed the CI Procedure, the corresponding action possesses moral worth. The difficulty, however, is that his four-step procedure admits counterexamples: a maxim may pass the test and yet, for Kant, still lack moral worth.

Consider the maxim “In order to obtain greater advantage, I will be honest and keep my promises.” Applied to the CI Procedure:

(1): In order to obtain greater advantage, I will be honest and keep my promises.

(2): In order to obtain greater advantage, everyone will be honest and keep their promises.

(3): In order to obtain greater advantage, everyone (as if by a law of nature) is honest and keeps their promises.

(4): We adjoin this “honesty-for-advantage” law of nature to the existing laws of nature and determine the resulting natural order.

No contradiction seems to arise. A world where everyone keeps promises to gain advantage would function smoothly; honesty would lower transaction costs and enhance reputation, which agents could rationally will. The maxim thus passes the CI Procedure and, under Rawls’s account, the action appears moral. Yet Kant would deny it moral worth, since the motive—advantage—is hypothetical.

This exposes a structural limitation. The CI Procedure tests only the form of maxims—their universalizability—not their motive. Many civic virtues (e.g., keeping promises for reputation) are universalizable but lack worth when done from inclination. Kant insists that “Act so that you use humanity, as much in your own person as in the person of every other, always at the same time as end and never merely as means.” [16] By this standard, honesty-for-advantage fails, since others are treated instrumentally.

The universalizability test discerns contradiction but not the ground of action. For Kant, moral worth depends on acting from duty. A procedure based solely on the Formula of Universal Law cannot examine this inner principle. To remedy this, the Formulas of Humanity and Autonomy must be integrated. A maxim must (1) be universalizable, (2) treat humanity as an end, and (3) express the autonomy of reason’s own legislation.

Rawls distinguishes these standpoints but limits his procedure to the first-person relation to the moral law. As he admits, “the categorical imperative procedure must not be merely formal but have sufficient structure to specify requirements on moral”. [17] Since his model is explicitly provisional, it allows extension.

Accordingly, the Formula of Humanity can serve as a further criterion within the procedure. I propose adding:

(5): In the adjusted order yielded at Step (4), humanity in the person of any individual is treated as an end, never merely as a means.

Re-testing “In order to obtain greater advantage, I will be honest,” Steps (1)–(4) reveal no contradiction, but Step (5) fails: the rational nature of others is used as a means to gain. Thus the maxim, though universalizable, lacks moral worth.

Yet the five-step version still cannot exclude heteronomous motives. Consider the maxim “I shall not kill,” derived from divine command. It passes the first five steps—it is universalizable, willable, and respects humanity—but fails Kant’s autonomy condition. As he writes, “Do no action in accordance with any other maxim, except one that could subsist with its being a universal law, and hence only so that the will could through its maxim at the same time consider itself as universally legislative.” [18] The authority of moral law derives not from external command but from reason’s self-legislation. Hence, a final step is required:

(6): The maxim exhibits the will’s principle of universal legislation—the law is autonomous, not heteronomous.

Re-testing the divine-command maxim: it satisfies Steps (1)–(5) but fails Step (6), for its source lies in divine authority, not autonomous reason. Any heteronomous maxim, even if universalizable and end-respecting, lacks moral worth.

The augmented CI Procedure therefore reads:

(1): I am to do X in circumstances C in order to bring about Y unless Z.

(2): Everyone is to do X in circumstances C in order to bring about Y unless Z.

(3): Everyone always does X in circumstances C in order to bring about Y, as if by a law of nature.

(4): We adjoin the as-if law of nature from Step (3) to the existing laws of nature and determine the resulting natural order.

(5): In the adjusted order yielded at Step (4), humanity in the person of any individual is treated as an end, never merely as a means.

(6): The maxim reflects the will’s principle of universal legislation: the law is autonomous, not heteronomous.

The transition from a four- to a six-step procedure can be viewed as three tiers: Tier 1 (Universality): whether the maxim can be a universal law and whether rational agents can will the resulting order. Tier 2 (End): whether the action treats humanity as an end in itself. Tier 3 (Legislation): whether the will is autonomous rather than heteronomous.

In sum, Rawls’s four-step procedure, grounded in the Formula of Universal Law, provides a clear mechanism for testing universality but cannot ensure moral worth. By integrating the Formulas of Humanity and Autonomy, the six-step augmentation preserves its structure while addressing its moral deficiency. It ensures that the CI Procedure yields genuinely correct moral judgments in Kant’s sense—those that are not only universalizable but also grounded in respect for humanity and in the self-legislation of reason.

5. Conclusion

Kant does not specify explicit steps for producing correct moral judgments, yet the Groundwork makes clear that the Categorical Imperative functions as the law of judgment itself, not as a mere corrective device. Wood misreads this relation by positing a temporal priority of intellectual capacity and treating moral principles as subsequent constraints. For Kant,

pure practical reason is prior, and the imperative directly generates correct moral judgments by testing maxims for moral worth.

Rawls, by contrast, reconstructs this process through the CI Procedure, in which moral judgment arises from testing whether a maxim can be a universal law. His interpretation accords more closely with Kant’s intent, though it remains incomplete: the four-step procedure derived from the Formula of Universal Law cannot assess the motive of action or the autonomy of the will. To address this, I propose extending the procedure by incorporating the Formulas of Humanity and Autonomy. The augmented six-step version adds an end-test—treating humanity as an end in itself—and a legislation-test—ensuring the will’s self-legislation. Together, they secure the moral worth of maxims and guarantee the generation of correct moral judgments in Kant’s sense.

Viewed within the broader realist–constructivist debate, both Rawls and Wood reinterpret Kant in light of their own commitments. Yet on the question of how correct moral judgments are produced, Rawls’s constructivist approach—once supplemented with the Humanity and Autonomy tests—offers a more coherent and textually grounded account than Wood’s realist correction model.

References

- [1] Wood, A. (2017). *Formulas of the Moral Law*. New York: Cambridge University Press, 21.
- [2] Kant. *Groundwork for the Metaphysics of Morals*. Edited and translated by Allen W. Wood. New Haven and London: Yale University Press, 19.
- [3] Kant. *Groundwork for the Metaphysics of Morals*. Edited and translated by Allen W. Wood. New Haven and London: Yale University Press, 37.
- [4] Wood, A. (2017). *Formulas of the Moral Law*. New York: Cambridge University Press, 22.
- [5] Wood, A. (2017). *Formulas of the Moral Law*. New York: Cambridge University Press, 39.
- [6] Wood, A. (2017). *Formulas of the Moral Law*. New York: Cambridge University Press, 25.
- [7] Kant. (2015). *Critique of Practical Reason*. Cambridge: Cambridge University Press, 41.
- [8] Kant. *Groundwork for the Metaphysics of Morals*. Edited and translated by Allen W. Wood. New Haven and London: Yale University Press, 54–55.
- [9] Kant. *Groundwork for the Metaphysics of Morals*. Edited and translated by Allen W. Wood. New Haven and London: Yale University Press, 41.
- [10] Rawls, J. (1999). *Collected Papers*. Edited by Samuel Freeman. London: Harvard University Press, 515.
- [11] Kant. *Groundwork for the Metaphysics of Morals*. Edited and translated by Allen W. Wood. New Haven and London: Yale University Press, 37.
- [12] Rawls, J. (1999). *Collected Papers*. Edited by Samuel Freeman. London: Harvard University Press, 498.
- [13] Rawls, J. (1999). *Collected Papers*. Edited by Samuel Freeman. London: Harvard University Press, 514.
- [14] Rawls, J. (2000). *Lectures on the History of Moral Philosophy*. Cambridge: Harvard University Press, 168–169.
- [15] Rawls, J. (2000). *Lectures on the History of Moral Philosophy*. Cambridge: Harvard University Press, 170.

- [16] Kant. *Groundwork for the Metaphysics of Morals*. Edited and translated by Allen W. Wood. New Haven and London: Yale University Press, 46-47.
- [17] Rawls, J. (2000). *Lectures on the History of Moral Philosophy*. Cambridge: Harvard University Press, 163.
- [18] Kant. *Groundwork for the Metaphysics of Morals*. Edited and translated by Allen W. Wood. New Haven and London: Yale University Press, 52.
- [19] Yang, Y. (2013). On Kant's influence on Rawls's theory of justice: On the significance of the history of philosophy for contemporary philosophical research. *Wuhan University Journal (Humanities and Social Sciences Edition)*, 3, 26-34.
- [20] Qiao, H. (2011). Rawls's "constructivist interpretation" of Kant's ethics: On the prototype of the justificatory structure in *A Theory of Justice*. *Philosophical Researches*, 12, 98-104.