Dialogue (2026), 1–22
doi:10.1017/S0012217326101012

CAMBRIDGE UNIVERSITY PRESS

ARTICLE

The Categorical Imperative Against Lying and Its Permissible Violations

Byeong D. Lee

Department of Philosophy, Sungkyunkwan University, Republic of Korea
Email: bydlee@skku.edu

Abstract

Immanuel Kant maintains that lying is always impermissible, since moral norms are categorical imperatives that admit of no exceptions. Yet this strict position conflicts with a widely shared intuition that lying may be permissible in certain exceptional cases. This article develops a revisionary Kantian proposal that accommodates this intuition while preserving the categorical character of moral norms. Central to this proposal is a default-and-exemption account of moral responsibility: The prohibition on lying applies universally, making agents morally responsible by default. This default condition, however, may be overridden if justifiable reasons for exemption exist.

Résumé

Emmanuel Kant soutient que le mensonge est toujours moralement inadmissible, les normes morales relevant d'impératifs catégoriques qui n'admettent aucune exception. Pourtant, cette position entre en tension avec l'intuition selon laquelle le mensonge peut être permis dans certains cas exceptionnels. Cet article présente une proposition kantienne révisionniste conciliant cette intuition avec le caractère catégorique des normes. Il défend une conception de la responsabilité fondée sur un schéma de « défaut et exemption » : l'interdiction de mentir vaut universellement, engageant par défaut la responsabilité des agents, mais pouvant être levée en présence de raisons justifiant l'exemption.

Keywords: categorical imperative; Kantian ethics; default condition of responsibility; exemption condition of responsibility; permissibility of lying

1. Introduction

According to Immanuel Kant, moral norms are categorical imperatives that bind all rational beings without exception. One such norm is that one ought not to lie. However, this view is challenged by the well-known Murderer-at-the-Door problem. Consider the familiar case of a would-be murderer who comes to the door of an

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Canadian Philosophical Association/L'Association Canadienne de Philosophie. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial licence (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use.

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

individual — call this person S — and asks about the whereabouts of S's friend. The murderer intends to kill the friend, and S knows where the friend is hiding. Thus, S must decide whether to lie in order to save the friend, or to tell the truth, thereby exposing the friend to mortal danger.

On Kant's view, lying is impermissible even in morally exigent circumstances such as this, because moral norms are categorical imperatives that admit of no exceptions. Yet this strict position conflicts with the widely shared intuition that lying may be morally permissible in certain exceptional cases. Moreover, if S tells the truth, he appears to violate another moral norm — namely, that one ought to help those in urgent need, if one can. This raises a pressing question: If both "one ought not to lie" and "one ought to help those in urgent need, if one can" are absolute moral norms, what ought one to do when they come into conflict?

This article proposes a novel resolution to the apparent tension between Kant's categorical prohibition on lying and the intuition that lying may, in certain exceptional circumstances, be morally permissible. This article proceeds as follows. Section 2 argues that the Murderer-at-the-Door case is best construed as involving a conflict between two moral norms. Section 3 argues against Kant's claim that moral norms cannot genuinely conflict. Section 4 outlines the default-and-exemption account of moral responsibility, which provides a crucial basis for my proposed resolution. Finally, drawing on this account, Section 5 argues that Kant's categorical imperative against lying can be reconciled with the possibility of its permissible violations.

One preliminary caveat is in order: This article does not aim to provide a novel exegetical interpretation of Kant's own position. Rather, it advances a broadly Kantian proposal that seeks to reconcile Kant's claim that lying is categorically prohibited with the intuition that, in rare cases, it may be morally permissible. The concern, then, is not whether this account preserves fidelity to Kant's doctrine, but whether it develops a version of Kantian ethics that accommodates both the unconditional authority of the categorical imperative and the permissibility of lying in exigent circumstances.

2. The Conflict of Moral Norms

According to Kant, the supreme principle of morality is the Categorical Imperative, which he articulates through three formulations: the Formula of Universal Law (Kant, 1996a, 4:421), the Formula of Humanity (Kant, 1996a, 4:429), and the Formula of Autonomy (Kant, 1996a, 4:431). The third formula is not directly relevant to the aim of this article. By contrast, the first two formulas bear directly on its topic and therefore merit brief explanation.

First, no practical norm qualifies as a moral norm unless it applies universally to all rational beings. Accordingly, we may say that a moral norm is formally correct if and only if it has the form of a universal law. Equivalently, for a maxim to qualify as a moral norm, it must be capable of being adopted as a universal norm without contradiction. Hence, the Formula of Universal Law can be understood as providing the formal correctness condition for moral norms.

Second, according to Kant, every rational being is an end in itself and so must be treated as such. Thus, we may say that a moral norm is materially correct if and

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

only if it conforms to the Formula of Humanity. Accordingly, the Formula of Humanity can be understood as providing the material correctness condition for moral norms.

What does it mean to treat a rational being as an end in itself? John Rawls (2000, pp. 190–195) distinguishes between two interpretations of the Formula of Humanity: a negative one and a positive one. Each rational being holds the equal status of an end in itself, thereby establishing limits that must not be transgressed. To treat a rational being as an end in itself in the negative sense is to refrain from any action that would violate these limits. This interpretation grounds duties of right, which impose stringent, non-discretionary obligations on us — for example, the duty not to kill another person. By contrast, to treat a rational being as an end in itself in the positive sense is to actively promote two obligatory ends: one's own perfection, and the happiness of others. This interpretation gives rise to duties of virtue, such as promoting the happiness of others. On Kant's view, although duties of virtue, no less than duties of right, are categorical imperatives, there is an important difference between them. Duties of right are narrow (perfect) duties that admit of no latitude in their fulfilment, whereas duties of virtue are wide (imperfect) duties that allow latitude, since agents must adopt certain obligatory ends but retain discretion concerning when, how, and to what extent these ends are to be pursued.

How, then, is the Categorical Imperative related to substantive moral norms (or categorical imperatives), such as "one ought not to kill another person" or "one ought not to lie"? As noted earlier, the Categorical Imperative is the supreme principle of morality and provides both the formal and the material correctness conditions for moral norms. A substantive practical norm therefore qualifies as a moral norm only insofar as it satisfies both conditions.

Now, with the above framework of Kantian ethics in mind, let us consider the following four claims:

(1) Not only duties of right, but also duties of virtue are categorical imperatives.
(2) A duty of virtue may be fulfilled only insofar as doing so does not violate any duty of right.
(3) There can be circumstances in which a duty of virtue admits of no latitude.
(4) There can be circumstances in which a duty of virtue that admits of no latitude can be fulfilled only if the agent violates a duty of right.

To begin with, claim (1) is unproblematic. Kant is explicit that both duties of right and duties of virtue are categorical imperatives. The difference between them does not concern their normative source or authority, but rather the manner of their obligation and enforcement. With regard to duties of right, no latitude is permitted, and violations of such duties may rightly be subject to negative sanctions, such as moral blame. By contrast, duties of virtue permit latitude — at least in ordinary cases — and therefore a failure to fulfill a duty of virtue does not make an agent liable to such sanctions.

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

In addition, claim (2) finds strong support in Kant's texts. Duties of right are narrow duties that admit of no latitude, whereas duties of virtue are wide duties that permit discretion in their fulfilment (Kant, 1996c, 6:390). The latitude characteristic of duties of virtue concerns when, how, and to what extent one fulfills them; yet this latitude must never be exercised in ways that conflict with the unconditional requirements of narrow duties. Accordingly, any such latitude must always be exercised in conformity with duties of right.

Moreover, although duties of virtue ordinarily admit of latitude, there are exceptional cases in which this latitude is severely constrained. Paradigmatic cases of urgent aid illustrate the point: When an agent is uniquely positioned to assist someone in dire need at little cost, refusal to do so rightly invites moral criticism. Admittedly, Kant does not explicitly claim that duties of virtue lose their latitude in such emergencies. Yet several prominent Kantian scholars — among them Barbara Herman (1993, 2022), Thomas Hill (2002), and Onora O'Neill (1989, 1996, 2013) — have defended, either explicitly or implicitly, the view that duties of virtue may, in certain circumstances, lose their characteristic latitude, especially in cases requiring urgent assistance. Moreover, apart from exegetical considerations, there is a strong moral rationale for this view. When an agent is uniquely positioned to aid another in dire need at minimal cost, failure to act plausibly amounts to a refusal to treat that person as an end in itself. The intuitive judgement that moral criticism is appropriate in such cases is exceedingly strong. For this reason, it is desirable for Kantian ethics to recognize that no latitude is permitted in such circumstances — that is, to endorse (3).

Finally, let us turn to (4), namely, that there can be circumstances in which a duty of virtue that admits of no latitude can be fulfilled only if the agent violates a duty of right. For example, consider a situation in which an agent is the only person who can help a person in urgent need without significant difficulty. In such a circumstance, no latitude is permitted, and failing to help may therefore be subject to moral blame. Yet this requirement can come into conflict with a duty of right. The Murderer-at-the-Door scenario illustrates precisely such a conflict.

Let us suppose that S is confronted by a would-be murderer who demands information about the whereabouts of S's friend. S recognizes the malicious intention underlying the inquiry and, as Kant puts it, finds himself in a situation in which he "cannot evade an answer of 'yes' or 'no'" (Kant, 1996d, 8:426). In these circumstances, S may reasonably judge that the only way to save his friend from imminent death is to provide false information, thereby preventing the murderer from carrying out his plan. S's friend is in grave and immediate peril, and S alone is in a position to provide effective assistance. Accordingly, the moral requirement to aid those in dire need applies in this case. Because this is an emergency, no latitude is permitted: Failing to aid the endangered friend would constitute a violation of the duty to assist those in desperate circumstances, and moral blame or criticism may rightly be directed at such a failure. In this case, moral criticism cannot be easily dismissed on the grounds that the relevant duty belongs to the category of duties of virtue. Under these conditions, the Murderer-at-the-Door case may be interpreted as one in which the obligation to

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

assist those in urgent need applies and thus as involving a genuine conflict between two moral norms: the prohibition against lying ("one ought not to lie"), and the obligation to aid those in desperate circumstances ("one ought to help a person in dire need, if one can").

Now note that the four claims cannot all be true together. Claim (3) asserts that there are circumstances in which a duty of virtue admits of no latitude; in such cases, the duty must be fulfilled in a determinate way. Given (1), this requirement has categorical force and thus binds unconditionally. Claim (4) maintains that, in some such circumstances, the only way to fulfill a duty of virtue that admits of no latitude is to violate a duty of right. Accordingly, under those conditions, the agent is required to perform an action that violates a duty of right. However, (2) imposes a strict constraint: Duties of virtue may be fulfilled only insofar as their fulfilment does not violate any duty of right. This constraint rules out precisely the kind of action that (4) identifies as required. Therefore, Kantian ethics must provide a way of resolving this inconsistency.[1]

3. Moral Norms as Social Norms

As argued in the previous section, the Murderer-at-the-Door case can be construed as a case in which two moral norms come into conflict. It is important to note, however, that Kant explicitly denies the possibility of such conflicts, maintaining that "a *collision of duties* and obligations is inconceivable" (Kant, 1996c, 6:224). In this section, I argue against this denial.

First, moral norms may be understood primarily as social norms. On Kant's view, however, morality centres on the capacity of rational agents to autonomously direct their own lives and is therefore not primarily concerned with the social regulation of individual conduct. I do not dispute this exegetical claim. Nevertheless, my aim is not to offer a novel interpretation of Kant's view but to articulate a more defensible version of Kantian ethics. From this revisionist standpoint, there are compelling reasons to regard moral norms as social norms. To do so is to understand them as governing relations among members of a moral community. On this view, moral norms are justified as intersubjective standards regulating relations among rational agents. Accordingly, these norms are ones that agents can demand not only of themselves but also of others. Within such a framework, social sanctions play an essential role. When a member of the moral community violates a moral norm — for example, the prohibition against killing — responses such as moral blame are justified. Without such practices, moral norms would lack the institutional and interpersonal normative force required to guide and coordinate conduct in a shared world. To claim

[1] There is considerable debate in the literature about the status of duties of right and their relationship to the Categorical Imperative (see, e.g., Guyer, 2002; Willaschek, 2002; Wood, 2002). While these exegetical disputes are illuminating, the aim of the present article is not to advance a novel interpretation of Kant, but rather to articulate a philosophically defensible Kantian framework that emphasizes the moral justification of duties of right as categorical imperatives. In particular, I assume that, insofar as duties of right are understood as categorical imperatives rather than merely as juridical requirements, their justification must ultimately be grounded in the supreme principle of morality.

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

that moral norms are primarily social norms is to affirm that they apply to agents as members of a moral community, regulating their interpersonal relations. Their effective operation depends on social forms — such as sanctioning practices — that help secure their normative authority: They rest on intersubjective justification, are upheld through positive and negative social sanctions, and are assessed intersubjectively with respect to compliance and violation. One further point worth noting is that the interpretation of moral norms as social norms aligns with Kant's idea of "a kingdom of ends" (Kant, 1996a, 4:433), whose relations are governed by universally valid norms. In this respect, moral norms have an essentially relational and public dimension: They regulate how agents are to stand in relation to one another under principles that all can share.

A cautionary note is necessary here: By claiming that moral norms are primarily social norms, I do not mean that their justification is reducible to existing social practices. Rather, their justification consists in the possibility of intersubjective rational endorsement among all agents. I will defend this claim in Section 5. At the same time, I do not deny Kant's point that we ought to perform morally right actions because they are morally right, rather than merely to avoid negative sanctions such as moral blame. This Kantian insight, however, should be distinguished from the conditions required to constitute a functioning moral community. Consider a case in which one person unjustifiably harms another but faces no negative social response. The absence of sanction neither nullifies the wrongness of the action nor releases others from their obligations. Yet systematic non-compliance, if left unchecked, undermines the practice of morality itself. When agents can no longer reasonably expect others to comply, the practical authority of moral norms as shared standards is weakened: Their ability to coordinate behaviour and guide mutual expectations is diminished. Consequently, although we must act morally because actions are right in themselves, the failure to impose negative social sanctions can jeopardize the realization of a moral community in which our relations are governed by moral norms rather than by the law of the jungle, where the strong prey on the weak. In this regard, Kant's point is instructive: A civil condition in which public right is secured can be achieved only by regulating our mutual relations through publicly lawful external coercion (Kant, 1996c, 6:312). Therefore, the social enforcement of moral norms ensures that they function as publicly shared standards, securing their practical authority without undermining their intrinsic moral justification.

Second, moral norms can hardly be imposed on us by identifying all possible exceptional cases in advance and codifying them as such, since it is not feasible to foresee all such cases. For a similar reason, Allen Wood writes:

> [I]n moral philosophy it is an important truth that because of the great complexities of human life, no moral rule simple enough to be practically useful can be framed so delicately as to be free of exceptions. (Wood, 2008, p. 250)

Hence, it is more reasonable to apply moral norms universally — even to cases that appear exceptional — and then assess whether there is reason to exempt agents from moral responsibility. I will elaborate on this point in the next section.

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

Third, moral norms function as social standards that regulate relationships among members of the moral community. To perform this role, they must be sufficiently general to apply across a wide range of circumstances. Such generality is crucial if they are to be publicly shared and readily understood by all members of the moral community. A norm that applies only to narrowly specified situations is unlikely to serve as a reliable guide amid the complexity and unpredictability of moral life. By contrast, general norms provide standards capable of guiding deliberation across diverse contexts and can be applied prospectively rather than retrospectively. Moreover, norms that lack sufficient generality cannot serve as common standards in moral deliberation or discussion. General norms enhance stability and enable mutual expectations: They allow individuals not only to anticipate their own actions but also to predict the actions of others, thereby facilitating coordination, cooperation, and trust. In short, if moral norms are to function as public standards that guide action, justify judgement, and regulate moral relations, they must possess a sufficient degree of generality.

Fourth, within Kantian ethics, there can be no hierarchy among moral norms or duties. Moral norms must be categorical imperatives, unconditionally obligatory for all rational beings as such. If, when a conflict arises between two norms, N_1 and N_2, N_1 is said to take precedence over N_2, this implies that N_2 is obligatory only on the condition that it does not conflict with N_1. In that case, N_2 would not be an unconditional categorical imperative. The claim that a hierarchy exists among moral norms is therefore incompatible with Kant's core commitment to their unconditional normative force.

A related point concerns the absence of any higher-order principle capable of resolving genuine conflicts among moral norms. Suppose that norms N_1 and N_2 come into conflict. For this to constitute a genuine conflict, both must qualify as genuine moral norms — that is, they must satisfy the formal and material correctness conditions. Whether a norm meets these conditions is ultimately determined by the supreme principle of morality. Accordingly, insofar as N_1 and N_2 both satisfy them, there can be no further principle that adjudicates conflicts between them. In this respect, it is noteworthy that both norms may be indispensable for treating rational beings as ends in themselves. For example, the norms that one ought not to lie and that one ought to help those in dire need, when one is able to do so, may each be necessary for fulfilling the requirement to respect humanity as an end in itself. Each norm therefore can be equally justified by the supreme principle of morality. It is true that, insofar as duties of virtue ordinarily admit of latitude, their fulfilment must remain within the bounds set by duties of right. Violating a duty of right does not fall within the scope of permissible latitude. In this qualified sense, duties of right may be said to take priority over duties of virtue. However, this priority holds only where latitude is in fact permitted. In exceptional cases in which duties of virtue admit no latitude, as argued in the previous section, duties of right do not take priority over such duties of virtue.

For these reasons, it is reasonable to recognize that genuine conflicts among moral norms can arise in certain exceptional cases. On this revisionary Kantian account, the

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

Murderer-at-the-Door case is best understood as involving a genuine conflict between two moral norms: the prohibition against lying, and the obligation to aid persons in desperate circumstances when the agent alone is able to do so.[2]

[2] Kant (1996c, 6:239, 383, 394) distinguishes duties of right from duties of virtue in terms of their external enforceability. Duties of right concern external actions that may infringe upon the equal freedom and rights of others; for this reason, they are strict, admit of no latitude, and are externally enforceable by juridical coercion. By contrast, duties of virtue allow for latitude on the part of the agent with respect to when, how, and to what extent they are to be fulfilled. While Kant draws this distinction in terms of external enforceability, the aim of the present article is not to offer a novel exegetical account of his position, but to articulate a version of Kantian ethics that is, in certain respects, more defensible. On the view I develop, the distinction between duties of right and duties of virtue is more adequately drawn not in terms of external enforceability, but in terms of two ways of treating rational beings as ends in themselves: a negative way and a positive way. Following Rawls, the negative interpretation of the Formula of Humanity grounds duties of right, which impose stringent, non-discretionary obligations, whereas the positive interpretation grounds duties of virtue, which allow for latitude in their fulfilment. The substantive upshot of this proposal is that duties of right are to be understood as moral requirements that impose stringent, non-discretionary obligations grounded in the negative requirement of the Formula of Humanity, thereby providing the moral basis for the justification of coercion, even if they are not themselves defined in terms of coercibility. On this account, the connection between right and coercion is preserved, but it is explanatory rather than definitional: Coercion is justified because such duties are violated, not because coercibility is built into their very concept.

The rationale for this approach is as follows. Duties of right are categorical imperatives, and the norms associated with them must be capable of functioning as social norms. To play this role, such norms must be publicly demandable: It must be possible to hold agents responsible for their violations through forms of social sanction, such as moral criticism. This notion of public demandability is broader than juridical coercion and does not coincide with external enforceability in Kant's juridical sense. Three considerations support this position.

First, external enforceability is not exempt from the demand for justification. A person subjected to juridical coercion for violating a moral norm may demand that such coercion be justified. To meet this demand, we must provide moral reasons that justify external enforceability in the juridical sense. Moreover, to avoid circularity, such reasons must be independent of external enforceability itself. In this respect, moral responsibility is more fundamental than juridical responsibility. On Kant's view, the supreme principle of morality is the Categorical Imperative. As pointed out in Section 2, the Formula of Humanity provides the material correctness conditions for moral norms. Accordingly, the moral reasons that justify external enforceability must be grounded in a failure to treat a person as an end in itself. Each person bears the status of an end in itself, thereby establishing limits that must not be transgressed. To treat a rational being as an end in itself, in this negative sense, is to refrain from any action that would violate these limits. External enforceability can, then, be justified insofar as it responds to violations of this requirement. On this interpretation, duties of right are defined independently of external enforceability while also supplying a more fundamental moral rationale for why such enforceability is justified.

Second, Kant classifies the duty not to commit suicide as a duty of virtue on the grounds that it is not externally enforceable. Yet there are strong reasons to regard this duty as a duty of right. The duty to refrain from suicide may be understood as having the same character as the duty to refrain from murder. Every human being, without exception, is an end in itself — and this includes oneself. To take one's own life without sufficient justification is therefore to fail to respect oneself as an end in itself. As Rawls (2000, p. 193) argues, the negative interpretation of the Formula of Humanity applies to suicide. Under this interpretation, each person's status as an end in itself establishes constraints that admit of no latitude. In this light, refraining from violating one's right to life can be understood as a stringent, non-discretionary requirement. At the same time, the fact that such a duty is not, as a matter of institutional fact, subject to juridical enforcement does not undermine its status as a duty of right on this account. Rather, it shows that juridical enforceability is not constitutive of duties of right, but contingent upon further considerations concerning the appropriate scope and limits of coercive institutions. Even in the absence of such enforcement, agents remain subject to moral demands grounded in the negative requirement of the Formula of Humanity.

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

4. The Default-and-Exemption Account of Moral Responsibility

As previously noted, Kant insists that the injunction "one ought not to lie" is a categorical imperative that admits of no exceptions. On his view, agents are required not to lie even in exigent circumstances, such as the Murderer-at-the-Door case. This uncompromising stance, however, conflicts with the widely shared intuition that lying may be morally permissible in exceptional cases of this kind. In what follows, I argue that Kant's fundamental claim — that "one ought not to lie" is a categorical imperative — can nonetheless be reconciled with this intuition. The key to this reconciliation lies in the default-and-exemption account of moral responsibility; this section therefore outlines that account.

As discussed in Section 2, the Formula of Universal Law provides the formal correctness condition for moral norms: A norm is formally correct if and only if it has the form of a universal law — that is, if and only if it can be applied without exception to all rational beings. Similarly, the Formula of Humanity provides the material correctness condition for moral norms: A norm is materially correct if and only if it conforms to the Formula of Humanity. Moreover, the Formula of Humanity admits both negative and positive interpretations. On the negative interpretation, we are required not to infringe upon the equal status of any rational being as an end in itself;

There is a further reason to classify the duty not to commit suicide in this way. Kant's principle of right requires the consistency of the freedom of each "with everyone's freedom in accordance with a universal law" (Kant, 1996c, 6:230). If it were morally permissible for each person to end their own life at will, the stability of a community of rational agents would be undermined. Accordingly, there is no need to restrict duties of right solely to the compatibility of one's actions with the freedom of others in a narrow juridical sense. The concept of right can instead be understood more broadly as encompassing fundamental constraints on the permissible use of freedom, including those grounded in one's own status as an end in itself.

Third, although duties of virtue may not be juridically coercible, they are not exempt from public moral criticism and demand. In emergency situations, an agent may be subject to moral blame for failing to assist a person in urgent danger when such assistance could be provided without significant difficulty. Such moral blame does not constitute juridical coercion, but it nonetheless exemplifies a form of public demandability characteristic of social norms.

For these reasons, distinguishing duties of right from duties of virtue on the basis of external enforceability in the juridical sense is problematic. The distinction is more adequately drawn in terms of two ways of treating rational beings as ends in themselves: negatively and positively. On this account, duties of right correspond to the negative requirement of the Formula of Humanity and impose strict, non-discretionary constraints, whereas duties of virtue correspond to its positive requirement and allow for latitude. This framework also explains why agents may be subject to moral blame for failing to assist others, even in the absence of juridical enforcement: Such cases reflect forms of public demandability rather than coercion.

I acknowledge that this revisionist approach diverges from Kant's own Doctrine of Right, in which external enforceability defines the nature of juridical duties. As scholars such as Helga Varden (2010) and Jacob Weinrib (2008) argue, Kant's arguments in "On a Supposed Right to Lie from Philanthropy" (Kant, 1996d) are framed from the juridical point of view. My proposal diverges from these exegetical interpretations by treating coercion as justified when certain stringent moral constraints are violated, rather than as built into the very definition of "right." In this respect, I do not deny the exegetical accuracy of their interpretations. Rather, I aim to articulate a revisionist Kantian framework in which the negative requirement of the Formula of Humanity provides the moral basis for coercion. This approach preserves the connection between right and coercion, but shifts the explanatory priority: Coercion is morally justified when duties of right — understood as strict, non-discretionary constraints — are transgressed. Thus, while Varden and Weinrib highlight the juridical point of view, my account seeks to broaden Kantian ethics by grounding right in the moral requirement to treat rational beings as ends in themselves.

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

on the positive interpretation, we are required to promote the happiness of rational beings.

What, then, is an example of a moral norm that satisfies these correctness conditions? One such example is the prohibition on killing another person. This norm satisfies both the formal and material correctness conditions. To kill another person is to violate their basic right to life, thereby infringing upon a duty of right grounded in the Formula of Humanity, construed negatively. For this reason, killing is morally impermissible, and the corresponding duty applies to all rational beings without exception. Nevertheless, there may be exceptional cases in which an agent acts contrary to such a norm without thereby being morally responsible. Consider, for instance, a case in which an agent S kills another person T in self-defence against T's attack. In such a case, S may be exempt from moral responsibility. Along these lines, it can be argued that if S kills T, the default condition is that S is morally responsible for T's death. However, if S acts in self-defence against T's aggression, S may be exempt by virtue of a justifiable reason for exemption.

Exemptions from moral responsibility may also admit of degrees. Suppose S kills T in self-defence, though T's threat is not genuinely life-threatening. In such a case, S's response may be excessive, making full exemption difficult. While S may bear reduced responsibility for T's death, complete exemption would be unjustified.

Furthermore, as noted in Section 3, moral norms can hardly be specified by identifying in advance all possible exceptional cases. The range of situations in which one might kill in self-defence is too diverse to be exhaustively anticipated. It is therefore more plausible to treat moral norms as universally applicable by default — even in cases that initially appear exceptional — and to determine, in light of particular circumstances, whether the default condition of responsibility is overridden.

These considerations support understanding moral responsibility in terms of the default-and-exemption account. On this account, when a rational agent violates a moral norm, the default condition is that the agent is morally responsible for the violation. This condition may, however, be overridden in exceptional cases in which the agent has a justifiable reason for exemption. In such cases, an agent who violates a moral norm may be wholly or partially exempt from responsibility. This constitutes the exemption condition.

Three merits of this account deserve particular attention.

First, it preserves Kant's fundamental commitment to the universality of moral norms while accommodating the intuition that agents may, in exceptional circumstances, be exempt from responsibility despite having violated them. The default condition explains why moral norms have the status of categorical imperatives: Whenever an agent violates such a norm, the presumption is that the agent is responsible, and it applies universally. At the same time, the exemption condition explains how agents may be exempt from responsibility in exceptional circumstances — namely, when there are overriding reasons for exemption. In this connection, it is crucial to distinguish between the universal authority of a moral norm and the blameworthiness of its violation. An analogy with epistemic norms is instructive here. Consider *modus ponens*, a basic rule of inference that all rational beings ought to follow. We do not regard young children or individuals with significant cognitive impairments as epistemically blameworthy for failing to conform to

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

it. This does not show that the rule fails to apply to them; rather, they may be exempt from epistemic blame for justifiable reasons. The same holds for moral norms. Cases in which it is inappropriate to hold an agent responsible for violating a moral norm do not entail that the norm itself is inapplicable.

Second, this account explains why agents who violate moral norms bear the burden of providing reasons for exemption. Unless agents — or their defenders — can offer justifiable reasons that override the default condition, that condition remains in force. For example, if agent A lies to agent B, A may be required to justify the action; in the absence of such justification, A remains subject to moral criticism. Exceptional cases are therefore best understood not as instances in which moral norms fail to apply, but as cases in which agents may be exempt from responsibility. If norms were genuinely inapplicable in such cases, it would be difficult to explain why violations continue to invite moral scrutiny or why agents are expected to offer justifications. The very practice of demanding such reasons presupposes that the relevant norms remain in force.

Third, this account allows that cases that appear to involve conflicts between moral norms — such as the Murderer-at-the-Door scenario — may constitute genuine moral conflicts. Even in such cases, the present account can explain which course of action is more reasonable by appealing to reasons for exemption. A more detailed account of this claim will be provided in the next section.

It is worth emphasizing that the default-and-exemption account is at least consistent with Kantian ethics and can therefore be incorporated within a Kantian framework. Kant himself frames moral responsibility in terms of imputability. "*Imputation* (*imputatio*) in the moral sense is the *judgment* by which someone is regarded as the author (*causa libera*) of an action, which is then called a *deed* (*factum*) and stands under laws" (Kant, 1996c, 6:227). To impute an action to an agent is thus to regard that agent as its author and to hold them responsible for it. Such a judgement implies that the action stands under laws — specifically, moral laws. Accordingly, insofar as an action accords with or violates a moral norm, the agent is subject to moral evaluation and is thereby assessable as responsible. The present account is consistent with this Kantian conception of imputability, while providing a more fully articulated framework for the attribution of responsibility. The norm "one ought not to lie" functions as an unconditional categorical imperative. When a rational agent violates such a norm, the imputation of responsibility constitutes the default condition. However, if there are justifiable reasons for exemption, the exemption condition requires that we either withdraw the imputation or assign a diminished degree of responsibility, depending on the nature of those reasons.[3]

There are also independent reasons to think that a Kantian framework can accommodate degrees of responsibility. Kant distinguishes between positive and negative freedom (Kant, 1996a, 4:446). Positive freedom may be understood in terms of the rational capacity to act in accordance with moral norms (see Allison, 2019, p. 455). On this view, moral norms apply to us without exception because we possess

[3] For a detailed discussion of Kant's account of imputation, see Claudia Blöser (2015), who further contends that Kant's practical philosophy can accommodate degrees of responsibility, albeit on distinct grounds.

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

this capacity; it would be pointless to impose moral requirements on beings that lack it altogether. Negative freedom, by contrast, can be understood in terms of the absence of determination by alien causes (see Kant, 1996a, 4:446). Anything outside the rational will can count as such an alien cause, including pathological desires and inclinations. We morally criticize certain actions in order to prevent their recurrence. However, in the case of actions that lie beyond an agent's rational control, demanding that the agent refrain from them is futile. For example, if a kleptomaniac steals something due to an irresistible urge, they may be regarded as lacking negative freedom with respect to this act. Negative freedom therefore requires not only the absence of illegitimate external constraints but also the absence of internal compulsion or other pathological factors. In other words, agents lack negative freedom with respect to their actions when those actions are determined by internal or external factors beyond their rational control. Along these lines, it can be argued that moral responsibility requires not only positive freedom but also negative freedom. Notably, whereas positive freedom is an all-or-nothing capacity, negative freedom admits of degrees. An agent may have had no rational control at all over the action they performed, may have lacked such control to a significant extent, or may have lacked it only partially.[4]

The above distinction between positive and negative freedom fits well with the default-and-exemption account. If an agent possesses positive freedom, the default condition of responsibility applies. However, insofar as the agent wholly or partially lacks negative freedom with respect to a particular action, the exemption condition may be invoked. Moreover, any moral theory that fails to accommodate degrees of responsibility is implausible. It is counterintuitive to attribute the same degree of responsibility to a child or a cognitively impaired individual as to a fully competent adult. Likewise, it is implausible to treat an action performed under a threat of death as equally blameworthy as one performed entirely voluntarily.

My account may be regarded as a version of Kantian ethics for three additional reasons. First, it accepts the supreme principle of morality, the Categorical Imperative, in its three formulations — the Formula of Universal Law, the Formula of Humanity, and the Formula of Autonomy — and interprets the first two as providing the formal and material correctness conditions for moral norms. Second, it endorses the Kantian view that moral evaluation depends not on consequences but on the reasons for which an agent acts. What matters are the agent's reasons — those that justify their actions to others. Since consequences often lie beyond our control, they are not directly relevant to moral assessment. Third, my account upholds the Kantian requirement of impartiality: Moral reasons must have the same normative force for all rational beings. If an agent has a justifiable reason for exemption in a given situation, then any agent in a relevantly similar situation must be equally exempt. Exemptions must therefore be universalizable. For these reasons, my account can be regarded as a version of Kantian ethics.[5]

[4] For a detailed discussion and defence of this Kantian account of freedom, see Lee (2024).

[5] For an extended treatment of the distinction between Kant's ethics and Kantian ethics, see O'Neill (1989, Preface and Part 3).

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

5. Categorical Imperatives and Its Permissible Violations

Drawing on the default-and-exemption account of moral responsibility articulated in the previous section, this section explains why lying may be permissible in certain exceptional circumstances, while preserving the Kantian claim that moral norms are categorical imperatives.

Consider again the Murderer-at-the-Door case. S faces only two options: either to lie in order to save his friend, or to disclose his friend's whereabouts, thereby enabling the murderer to kill him. Both alternatives involve moral wrongdoing. The first violates the prohibition against lying; the second violates the duty to aid those in dire need. Thus, each option entails a violation of a moral norm, yet S must choose between them. The central question, then, is which option S ought to choose. In circumstances of this kind, the best S can do is to choose the morally less bad option available. Which option, then, is morally less bad?

The first option — lying — fails to accord with the Formula of Humanity. Let me explain. According to Kant, lying consists in making an untruthful statement with the intention of deceiving another into believing it to be true. He calls lying in this sense lying "in the ethical sense" (Kant, 1996c, 6:430). In his view, lying in the ethical sense requires that a statement be made, and that the statement be believed to be false by the person who makes it.[6] It is not required, however, that the statement be false. Thus, a lie in the ethical sense (henceforth, simply a lie) may turn out to be true. In addition, the moral status of a lie does not depend on whether the deception succeeds; even if no one is misled, the act still counts as a lie in the ethical sense. Moreover, some lies can infringe upon another person's juridical rights. For instance, if one falsely claims that a contract has been legally concluded when it has not, thereby causing legal harm to another party, this constitutes a violation of that party's juridical rights. Yet many lies bear no relation to the violation of juridical rights. Thus, the mere fact that certain lies can undermine juridical rights does not explain why all lies are morally impermissible. A deeper question, then, concerns the justification of the moral norm "one ought not to lie." What grounds the claim that every lie is morally impermissible? This question can be addressed by appeal to Kant's Formula of Humanity.

Rational agents must be free to set their own permissible ends, to determine the means appropriate to their realization, and to act upon those determinations, provided that these remain compatible with the equal freedom of others. Lying violates this requirement, for it constitutes an intentional attempt to distort another person's understanding of reality, thereby undermining their capacity to make rational choices on the basis of accurate information. What is decisive here is not whether this impairment succeeds, but rather that the maxim of lying subordinates another's

[6] According to Kant, it is possible to lie to oneself (Kant, 1996c, 6:429). For A's telling B that p to count as a lie, however, A must believe p to be false and must intend that B believe p to be true. If this intention succeeds, B will come to believe that p is true. On this account, a lie to oneself would be a case in which A and B are identical. One would therefore have to bring it about that one believes p while at the same time not believing p. Whether such self-deception is genuinely possible remains philosophically contentious. Even if it is possible, it may still undermine the principle of treating oneself as an end in itself, insofar as it compromises one's capacity to make rational decisions on the basis of accurate information. Moreover, the kind of lying at issue in the Murderer-at-the-Door case — the focus of this article — is lying to others. I shall therefore restrict the discussion to that form of deception.

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

rational agency to one's own ends. To interfere with another's autonomy in this way is to treat them as a mere means rather than as an end in itself.

Moreover, given our epistemic limitations, we are necessarily dependent on others for information — a dependence that underwrites the indispensable role of linguistic communication. The point is not merely that, if lying were generally permitted, the trustworthiness of testimony would collapse. Rather, the practice of testimony itself presupposes a norm of truthfulness, such that a maxim permitting lying cannot be coherently willed as a universal condition of shared epistemic agency. As a result, each person's capacity to choose appropriate means to their ends on the basis of true information would be undermined. In this context, Kant writes:

> Whoever may have told me a lie, I do him no wrong if I lie to him in return, but I violate the right of mankind; for I have acted contrary to the condition, and the means, under which a society of men can come about, and thus contrary to the right of humanity. [. . .] If a man publishes a false report, he thereby does no wrong to anyone in particular, but offends against mankind, for if that were to become general, the human craving for knowledge would be thwarted; apart from speculation, I have only two ways of enlarging my store of information: by experience, and by testimony. But now since I cannot experience everything myself, if the reports of others were to be false tidings, the desire for knowledge could not be satisfied. (Kant, 1997, 27:447–448)

Along these lines, Kant argues that truthfulness is "a human being's duty to everyone," and that every lie is "a wrong inflicted upon humanity generally" (Kant, 1996d, 8:426). To put the point another way, every lie is morally impermissible because it violates at least a "right of mankind," or a "right of humanity," understood as a non-juridical right grounded in the conditions of mutual rational interaction. Therefore, if S in the Murderer-at-the-Door case takes the first option, he not only infringes upon the would-be murderer's capacity for rational decision-making but also violates a right of humanity.[7]

Let us now turn to the second option. Disclosing the required information would foreseeably result in the death of an innocent person — a grave and irreversible harm. In addition, providing information that enables the commission of a wrongful act plausibly counts as assisting that act, and assisting serious wrongdoing is itself morally impermissible. Accordingly, one ought not to provide information required for the commission of murder. In such circumstances, the only way to avoid providing such assistance may be to lie — for example, by falsely claiming ignorance.

If these considerations are correct, S — or a defender of S's action — may argue that lying is the morally less bad option in the Murderer-at-the-Door case. By assumption, S must choose either to disclose his friend's whereabouts to a would-be

[7] It is important to note that, in the ethical sense, every lie can be understood as a violation of a duty of right. As argued in Section 2, the Formula of Humanity admits of both negative and positive interpretations. When construed negatively, it yields duties of right. On this reading, undermining another's capacity to make rational decisions on the basis of accurate information is not merely a failure to promote that person's happiness, but a direct infringement of their status as an end in itself — establishing limits that must not be transgressed.

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

murderer or to lie. In the former case, his friend is likely to be killed. From the standpoint of each person's status as an end in itself, the most serious infringement one can suffer is being killed by another. In other words, killing a person is the most serious violation of their status as an end in itself, as it inflicts an irreversible and profound harm. Moreover, disclosing the requested information would amount to assisting in serious wrongdoing. In the latter case, by contrast, S's friend is likely to be saved. Admittedly, this option involves interfering with the murderer's rational decision-making. However, interfering with the rational deliberation of an agent who intends to commit murder, though a genuine wrong, is less severe. It is rare for a single instance of misinformation to undermine a person's life as a whole. Consequently, deliberately providing false information that undermines a person's capacity to make a particular rational decision on the basis of truth is, in most cases, a far less serious infringement. Moreover, exempting agents from moral responsibility for lying in highly exceptional cases would not plausibly undermine the general trustworthiness of linguistic communication. In light of these considerations, it is difficult to maintain that lying is morally worse than allowing the murder of an innocent person. Accordingly, the fact that lying is unavoidable in order to prevent complicity in serious wrongdoing constitutes a compelling reason for exemption. One additional point worth mentioning is that, as noted in the previous section, a Kantian framework can accommodate degrees of responsibility. Along these lines, S — or a defender of S's action — may argue that lying is the morally less bad option in this case.

If this line of argument is sound, S cannot be criticized on the grounds that he ought to have chosen a better alternative. To hold S morally blameworthy, one must be able to maintain that he ought to have acted otherwise. Yet it is difficult to argue that allowing the murder would have been the morally better option. Consequently, even if S lies, he cannot be faulted for failing to make a better choice. In such a case, S possesses a justifiable reason for exemption: There is no better alternative to saving his friend's life. This reason overrides the default condition of responsibility for lying. Accordingly, S may be exempt from moral responsibility for his lie. To say that S is exempt from responsibility for lying is, in turn, tantamount to saying that the lie is morally permissible. The default-and-exemption account of responsibility thus allows us both to preserve the Kantian view that the norm "one ought not to lie" is a categorical imperative and to acknowledge that lying may be permissible in exceptional cases such as the Murderer-at-the-Door.

It is worth emphasizing that this account is not a form of consequentialism. As noted in the previous section, it affirms the core theses of Kant's moral philosophy. While there may be exceptional cases in which an agent is exempt from responsibility despite violating a moral norm, such norms retain their categorical normative force for all persons. The fact that determining whether the exemption condition is satisfied may involve comparing options to determine which is less morally bad does not entail a consequentialist evaluation of the relative priority of moral norms. Such comparisons do not determine which moral norm takes precedence; rather, they provide reasons for holding that, in rare circumstances of genuine moral conflict, agents may be exempt from responsibility even though they have violated a moral norm. Moreover, whether the exemption condition is satisfied in a given case does not depend on the actual

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

consequences of the action. What matters instead is whether, at the moment of decision, it was reasonable for the agent to believe that the exemption condition obtains.

At this point, one might wonder whether my account can provide a universal, public, and objective criterion for determining the circumstances under which an exemption is appropriate. A Sellarsian coherence theory of justification offers the resources to address this concern.

The coherentist insight is that we have no rational alternative but to evaluate the truth or justification of a belief within our conceptual framework. Two points deserve emphasis. First, as Ludwig Wittgenstein (1969, §§341–343) observes, genuine doubt is possible only against a background of beliefs that are not simultaneously doubted. Second, as Kant insists, reason is the highest court of appeal for all questions concerning itself and its principles. Accordingly, reason must answer such questions from its own resources.[8] Put differently, our conceptual framework provides the norms, criteria, and rules for defending or criticizing any claim, and so it is inevitable that demands for justification be addressed within that framework. Here, a "conceptual framework" may be understood as a framework within which genuine doubt is settled and demands for justification are met.

We may further adopt a Lehrerian account of coherence. Coherence is often characterized in terms of how the elements of a belief system fit together, but beyond such metaphorical descriptions, no generally accepted definition is available. In response to this difficulty, Keith Lehrer offers an illuminating proposal: Coherence can be understood in terms of "answering all objections" or "beating all competitors" (Lehrer, 2000, p. 129). One's belief that p coheres with one's background system just in case one can answer all objections raised against it on the basis of that system. However, as Wilfrid Sellars (1963) argues, the very concept of justification arises from our social practices of giving and asking for reasons. Lehrer's account should therefore be recast in intersubjective terms. On the Sellarsian version of coherence theory that I defend, we are justified in believing that p just in case we can answer all objections raised against it within our social practice of demanding justification and responding to such demands. Although this account has primarily been developed for epistemic justification, the underlying insight applies equally to practical justification: When we are required to justify actions or practical norms, we likewise have no rational alternative but to do so within our social practice of justification.[9]

With this social practice model of justification in place, we can see how a Kantian framework accommodates intersubjective assessments of the exemption condition of responsibility. As argued in Section 4, the default-and-exemption account is at least consistent with Kantian ethics and can therefore be incorporated into such a framework. In particular, my account upholds the Kantian requirement of impartiality: Moral reasons must have the same normative force for all rational beings. If an agent has a justifiable reason for exemption in a given situation, then any agent in a relevantly similar situation must be equally exempt. Exemptions must therefore be universalizable. Accordingly, insofar as it is successfully established within our social

[8] See, e.g., Kant (1998, A xi-xii, A 477/B 505) and Kant (1996b, 5:16).

[9] For a detailed discussion and defence of this Sellarsian social practice theory of justification, see Lee (2017, 2021).

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

practice of justification that a reason offered in support of an exemption is universalizable, that reason qualifies as a justifiable reason for exemption. Therefore, in the Murderer-at-the-Door case under consideration, insofar as S — or a defender of S's lie — successfully argues that he has a jusifiable reason that exempts him from responsibility for the lie, S is not to be held responsible for it.[10]

Finally, to avoid potential confusion regarding my claim that lying may be "permissible" in certain exceptional cases, we may distinguish between two senses of permissibility: strict norm-relative permissibility, and responsibility-relative permissibility. This distinction may be clarified as follows. For Kant, permissibility is not a comparative notion grounded in the absence of better alternatives; rather, it is a categorical status determined solely by conformity with the moral law. An action that violates a moral norm remains impermissible, regardless of whether all available options are bad. Moreover, Kant distinguishes between the objective moral status of an action and the agent's culpability. The fact that an agent is not morally blameworthy does not entail that the action itself is permissible. He allows that agents may be excused — or at least partially exculpated — on account of ignorance, coercion, or severe limitations on their capacity to act, while maintaining that the action remains contrary to duty.

I do not, however, accept Kant's claim that genuine conflicts of duty are impossible. Any plausible moral theory must acknowledge tragic cases in which wrongdoing is unavoidable. To address such cases, I adopt the default-and-exemption account of moral responsibility. On this account, when a rational being violates a moral norm, the default condition is that they are morally responsible for the violation. This default

[10] The Sellarsian social practice theory of justification can accommodate objectivity as well. As Robert Brandom (1994) argues, objectivity depends on distinguishing what is objectively correct from what is merely taken to be correct. On this view, objectivity is attitude-transcendental in the sense that anyone may be mistaken about what is objectively correct, and it is best understood as a feature of discursive intersubjectivity. Donald Davidson (1984) similarly emphasizes that the very idea of a genuinely alternative conceptual scheme is unintelligible within rational discourse. Suppose that a belief that p is genuinely justified within our social practice of justification. If a contradictory belief, $\neg p$, were also genuinely justified within some purportedly alien framework, there would have to be a reason in favour of $\neg p$. If that consideration is intelligible as a reason, then it is, in principle, available for assessment within our justificatory practice and would thereby undermine our justification for p. If, by contrast, it is unintelligible as a reason for us, then it cannot undermine our justification for p. Thus, the very notion of a genuinely alternative conceptual scheme cannot be made intelligible within rational discourse. Any demand for justification must be met within the space of reasons constituted by our social practice. Crucially, this practice is governed by a regulative ideal of unrestricted participation: In principle, no rational agent is excluded. If any rational agent were to present sufficiently strong countervailing reasons, we would be required — on grounds of rationality — to revise or abandon our beliefs. Accordingly, there are no membership restrictions on our social practice of justification; any rational being could, in principle, participate in it. From this perspective, a belief is genuinely justified not because it is currently accepted, but because it retains its justificatory status under ongoing rational scrutiny. If a belief's justificatory status does not depend on any undefeated error — such that no future rational criticism would overturn it — then it is really justified. Put differently, if a claim is justified within our social practice and remains undefeated in that ongoing, dynamic practice, it may be regarded as objectively true. For example, the proposition that the earth is round is objectively true: Although fallible, its justification is sufficiently robust that future inquiry is unlikely to overturn it. Therefore, fallibility is compatible with objectivity. The same general point extends to practical justification, though a full defence lies beyond the scope of this article. For a detailed discussion and defence of this account of objectivity, see Lee (2017, 2022).

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

condition can be overridden in exceptional cases where the agent has a justifiable reason for exemption. In other words, an agent who violates a moral norm for such a reason may be — at least partially — exempt from responsibility. This constitutes the exemption condition of moral responsibility. Within this framework, we may distinguish between two senses of permissibility. The first is strict norm-relative permissibility: An action is permissible only if it accords with moral norms. In such permissible cases, the agent bears no moral responsibility. This corresponds to the standard understanding of permissibility, according to which to say that an action is permissible is simply to say that it is not morally wrong. The second is responsibility-relative permissibility: An action may count as permissible under exceptional circumstances in which no better option is available and the agent is exempt from responsibility for performing it. In this sense, an action may be practically permissible — that is, one may perform it without being subject to moral blame, even though it remains normatively wrong.

If responsibility is the primary locus of moral appraisal in practice, it is natural to introduce a derivative evaluative category for actions that, while violating norms, are nonetheless exempt from blame. Our moral practice implicitly relies on this responsibility-sensitive notion of permissibility, especially in the context of practical deliberation. Consider an agent who is faced with two morally bad options: In such a case, we may appeal to the considerations that it is rational to choose the lesser evil, together with the possibility that the agent is exempt from responsibility for doing so. On this basis, we can legitimately say: "You may choose the morally less bad option." To recommend such a choice is, in effect, to confer permissibility upon it. Moral advice under tragic conditions thus presupposes a conception of permissibility tied to responsibility and justification rather than strict norm-conformity. To recommend an action in such circumstances is, practically speaking, to permit it. Any adequate moral theory should be able to accommodate this intuitive judgement.

In short, we can distinguish two levels of moral evaluation: one assessing conformity to moral norms, and the other assessing responsibility in concrete circumstances. The default-and-exemption account allows us to maintain that moral duties remain strict and exceptionless at the level of norms, while responsibility remains sensitive to the agent's practical situation — including cases in which every available option involves a violation of a moral norm. This distinction enables us to separate strict normative permissibility from responsibility-relative permissibility: that is, what is permissible at the level of moral norms from what is permissible at the level of practical deliberation. Crucially, allowing less bad choices does not implicitly rank moral norms or duties. The comparison operates at the level of practical decision-making under conditions of norm violation, not at the level of the norms themselves. Moral norms remain strict and exceptionless *qua* norms. By distinguishing between these two senses of permissibility, my proposal preserves Kant's strict view of the categorical authority of moral norms while offering a more realistic account of responsibility in exceptional cases of moral conflict.[11]

[11] One question worth addressing at this point is whether, in the Murderer-at-the-Door case, it is permissible for S to answer the would-be murderer's question truthfully. By assumption, if S answers truthfully, the default condition is that he bears responsibility for violating the moral norm requiring aid to

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

6. Conclusion

On Kant's view, no practical norm qualifies as a moral norm unless it applies universally to all rational beings, and the moral prohibition on lying is no exception to this requirement. This demand for universal applicability, however, gives rise to the Murderer-at-the-Door problem. Many share the intuition that lying may be morally permissible in exceptional cases of this kind, and Kantian ethics must therefore explain how such an intuition can be accommodated within its framework.

This article has argued that the categorical prohibition against lying can be reconciled with the possibility of its permissible violations by appealing to the default-and-exemption account of moral responsibility. On this account, when a rational agent violates a moral norm, the default condition is that the agent is morally responsible for the violation. This condition, however, may be overridden in exceptional circumstances where justifiable reasons for exemption exist. When such reasons obtain, an agent may be exempt, at least to some extent, from moral responsibility.

In the Murderer-at-the-Door case, if S lies in order to save his friend, he thereby violates the moral norm against lying. Accordingly, the default condition applies, thereby making S morally responsible for the lie. In this scenario, however, a justifiable reason for exemption may plausibly override the default condition. S can reasonably claim that lying was morally less bad than allowing an innocent person to be killed. Unless this claim is successfully challenged within our social practice of justification, S may be exempt from responsibility for his lie on the grounds that there was no morally better alternative.

Taken together, the moral norm that one ought not to lie applies universally to all rational agents, while exceptional cases such as the Murderer-at-the-Door illustrate how justifiable reasons for exemption can override the default condition of responsibility. Accordingly, it may be practically permissible for S to lie in order to save his friend.

Finally, it is worth briefly highlighting the most distinctive feature of my proposal. While I argue that Kantian ethics can accommodate the widely shared intuition that lying may be permissible in exceptional cases, my proposal diverges from prominent Kantian interpretations — such as those articulated by Herbert Paton (1954), Roger Sullivan (1994), Christine Korsgaard (1996), and Allen Wood (2008) — which likewise aim to accommodate this intuition. Although these approaches differ significantly in their specific proposals, they converge on a central thesis: Paradigmatic cases such as the Murderer-at-the-Door, in which moral duties appear to conflict, can

persons in imminent danger. The question, then, is whether there are reasons for exemption sufficient to override this default condition. In this context, S may face the criticism that he ought to have chosen differently, and it is difficult to see how such criticism could be fully addressed. One possible defence is that, when moral norms collide, there is no universal principle for determining which option constitutes the lesser wrong. In such circumstances, agents may struggle to judge correctly, and ignorance may partially excuse their choice. Yet this excuse cannot entirely remove moral blame, since even in exceptional cases, it remains rational to choose, as far as possible, the option that is morally less bad. For this reason, it is difficult to maintain that S's truthful answer to the murderer is wholly permissible. At best, it is a choice made under moral uncertainty — one that still leaves room for the criticism that a better decision should have been made.

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

ultimately be construed as not involving a genuine conflict of duties. Sullivan, for example, contends that what conflicts in such cases are not duties themselves but the grounds that ordinarily give rise to them, invoking Kant's distinction between duties and the grounds of obligation. Korsgaard, by contrast, maintains that lying may be permissible under certain conditions, insofar as a maxim such as "I will lie to deceivers when doing so thwarts the intended effects of their deception" can be willed as a universal law. On her account, the permissibility of lying is thus explained by the applicability, in exceptional cases, of a universalizable maxim permitting it.[12]

By contrast, my proposal rejects the widely shared assumption that apparent cases of duty conflict must be reinterpreted so as to avoid genuine moral conflict. Instead, it treats exceptional cases such as the Murderer-at-the-Door as involving a real conflict of moral norms. As argued in Section 3, there are good reasons to acknowledge the possibility of genuine conflicts among moral norms. Accordingly, the Murderer-at-the-Door case is best understood as involving a genuine conflict between two moral norms: the norm prohibiting lying, and the norm requiring assistance to a person in dire need. The philosophical significance of such cases lies precisely in the fact that every available option involves moral wrongdoing. Attempts to resolve the problem by denying the possibility of genuine conflicts between moral norms therefore amount to an evasion of these dilemmas rather than a serious engagement with them.

Admittedly, some theorists explain the permissibility of lying by accepting the possibility of genuine conflicts among moral duties. Most notably, Thomas Carson (2010) argues that the duty not to lie is a *prima facie* duty that can conflict with, and be overridden by, other moral duties, such as the duty not to harm others. In doing so, he explicitly rejects the Kantian absolutist thesis that the duty not to lie is exceptionless and incapable of being overridden. In this respect, Carson's view aligns closely with W. D. Ross's (1930) pluralism, according to which multiple *prima facie* duties — such as fidelity, non-maleficence, and beneficence — may come into conflict, and moral deliberation consists in weighing them against one another. This diverges from Kant's monistic framework, within which categorical imperatives are grounded in a single supreme principle of morality and do not admit of being overridden by other categorical imperatives.

One central merit of my proposal is that it preserves Kant's strict view that the moral norm prohibiting lying is a universal norm that applies without exception and cannot be overridden by other moral norms, even in highly exceptional circumstances. On this account, moral norms do not admit of exceptions or subordination to competing norms. The appearance of exceptional permissibility is instead explained at the level of moral responsibility: Although lying always violates the moral norm

[12] A cautionary note is in order. Korsgaard arrives at different conclusions regarding the permissibility of lying depending on which formulation of the Categorical Imperative is employed. On her view, lying to a would-be murderer may be permissible under the Formula of Universal Law, but impermissible under the Formula of Humanity. Her position thus implies that the two formulas are not equivalent (see Korsgaard 1996, pp. 134–135, 143). As noted in Section 2, however, the Formula of Universal Law can be construed as providing the formal correctness condition for moral norms, while the Formula of Humanity can be construed as providing the material correctness condition. Accordingly, a maxim qualifies as a moral norm only if it satisfies both formulations of the Categorical Imperative. If this is correct, the two formulas should be regarded as jointly necessary in determining the moral status of any given maxim.

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

against lying, there may be exceptional cases in which agents have justifiable reasons sufficient to exempt them from moral blame. By locating flexibility in our practice of responsibility assessment rather than in the categorical force of moral norms themselves, my proposal accommodates the intuition that lying may be permissible in exceptional cases without abandoning Kantian absolutism.

In conclusion, the Kantian proposal developed in this article offers a more compelling reconciliation of the tension between Kant's fundamental claim that the moral norm that one ought not to lie is a categorical imperative and the widely shared intuition that lying may nevertheless be morally permissible in certain exceptional cases.

Competing interests. The author declares none.

References

Allison, H. E. (2019). *Kant's conception of freedom: A developmental and critical analysis*. Cambridge University Press. https://doi.org/10.1017/9781316535196

Blöser, C. (2015). Degrees of responsibility in Kant's practical philosophy. *Kantian Review*, *20*(2), 183–209. https://doi.org/10.1017/S1369415415000011

Brandom, R. B. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard University Press. https://www.hup.harvard.edu/books/9780674543300

Carson, T. L. (2010). *Lying and deception: Theory and practice*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199577415.001.0001

Davidson, D. (1984). On the very idea of a conceptual scheme. In his *Inquiries into truth and interpretation* (pp. 183–198). Clarendon Press.

Guyer, P. (2002). Kant's deductions of the principles of right. In M. Timmons (Ed.), *Kant's metaphysics of morals: Interpretative essays* (pp. 23–64). Oxford University Press. https://global.oup.com/academic/product/kants-metaphysics-of-morals-9780198250104?cc=ca&lang=en&

Herman, B. (1993). *The practice of moral judgment*. Harvard University Press. https://www.hup.harvard.edu/books/9780674697188

Herman, B. (2022). *Kantian commitments: Essays on moral theory and practice*. Oxford University Press. https://global.oup.com/academic/product/kantian-commitments-9780192844965?cc=ca&lang=en&

Hill, T. E., Jr. (2002). *Human welfare and moral worth: Kantian perspectives*. Oxford University Press. https://doi.org/10.1093/0199252637.001.0001

Kant, I. (1996a). Groundwork of the metaphysics of morals (1785). In M. J. Gregor (Ed. & Trans.), *Practical philosophy* (pp. 37–108). Cambridge University Press. https://doi.org/10.1017/CBO9780511813306.007

Kant, I. (1996b). *Critique of practical reason* (1788). In M. J. Gregor (Ed. & Trans.), *Practical philosophy* (pp. 133–272). Cambridge University Press. https://doi.org/10.1017/CBO9780511813306.010

Kant, I. (1996c). The metaphysics of morals (1797). In M. J. Gregor (Ed. & Trans.), *Practical philosophy* (pp. 353–604). Cambridge University Press. https://doi.org/10.1017/CBO9780511813306.013

Kant, I. (1996d). On a supposed right to lie from philanthropy (1797). In M. J. Gregor (Ed. & Trans.), *Practical philosophy* (pp. 605–616). Cambridge University Press. https://doi.org/10.1017/CBO9780511813306.014

Kant, I. (1997). *Lectures on ethics* (P. Heath, Trans. & Ed., & J. B. Schneewind, Ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781107049512

Kant, I. (1998). *Critique of pure reason* (P. Guyer & A. Wood, Eds. & Trans.). Cambridge University Press. https://doi.org/10.1017/CBO9780511804649

Korsgaard, C. M. (1996). *Creating the kingdom of ends*. Cambridge University Press. https://doi.org/10.1017/CBO9781139174503

Lee, B. D. (2017). The truth-conduciveness problem of coherentism and a Sellarsian explanatory coherence theory. *International Journal of Philosophical Studies*, *25*(1), 63–79. https://doi.org/10.1080/09672559.2016.1236140

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press

Lee, B. D. (2021). A coherentist justification of epistemic principles and its merits. *Acta Analytica*, *36*(4), 533–551. https://doi.org/10.1007/s12136-021-00463-3

Lee, B. D. (2022). Knowledge as objectively justified belief. *Acta Analytica*, *37*(3), 397–414. https://doi.org/10.1007/s12136-021-00495-9

Lee, B. D. (2024). The manipulation argument and a Kantian account of freedom. *Theoria*, *90*(6), 567–582. https://doi.org/10.1111/theo.12556

Lehrer, K. (2000). *Theory of knowledge* (2nd ed.). Westview. https://www.routledge.com/Theory-Of-Knowledge-Second-Edition/Lehrer/p/book/9780813390536

O'Neill, O. (1989). *Constructions of reason: Explorations of Kant's practical philosophy*. Cambridge University Press. https://doi.org/10.1017/CBO9781139173773

O'Neill, O. (1996). *Towards justice and virtue: A constructive account of practical reasoning*. Cambridge University Press. https://doi.org/10.1017/CBO9780511621239

O'Neill, O. (2013). *Acting on principle: An essay on Kantian ethics* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139565097

Paton, H. J. (1954). An alleged right to lie. A problem in Kantian ethics. *Kant-Studien*, *45*(1–4), 190–203. https://doi.org/10.1515/kant.1954.45.1-4.190

Rawls, J. (2000). *Lectures on the history of moral philosophy* (B. Herman, Ed.). Harvard University Press. https://www.hup.harvard.edu/books/9780674004429

Ross, W. D. (1930). *The right and the good*. Clarendon Press. https://spot.colorado.edu/~heathwoo/readings/ross.pdf

Sellars, W. (1963). *Science, perception and reality*. Ridgeview Publishing Company. https://sites.pitt.edu/~rbrandom/Courses/2023%20Sellars/Sellars%20texts/Sellars-Wilfrid-Science-Perception-and-reality.pdf

Sullivan, R. J. (1994). *An introduction to Kant's ethics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511606151

Varden, H. (2010). Kant and lying to the murderer at the door … One more time: Kant's legal philosophy and lies to murderers and Nazis. *Journal of Social Philosophy*, *41*(4), 403–421. https://doi.org/10.1111/j.1467-9833.2010.01507.x

Weinrib, J. (2008). The juridical significance of Kant's "supposed right to lie." *Kantian Review*, *13*(1), 141–170. https://doi.org/10.1017/S1369415400001126

Willaschek, M. (2002). Which imperatives for right? On the non-prescriptive character of juridical laws in Kant's *Metaphysics of morals*. In M. Timmons (Ed.), *Kant's metaphysics of morals: Interpretative essays* (pp. 65–87). Oxford University Press. https://global.oup.com/academic/product/kants-metaphysics-of-morals-9780198250104?cc=ca&lang=en&

Wittgenstein, L. (1969). *On certainty* (G. E. M. Anscombe & G. H. von Wright, Eds., & D. Paul & G. E. M. Anscombe (Trans.). Harper & Row. https://www.harpercollins.com/products/on-certainty-ludwig-wittgenstein?variant=32154010517538

Wood, A. (2002). The final form of Kant's practical philosophy. In M. Timmons (Ed.), *Kant's metaphysics of morals: Interpretative essays* (pp. 1–21). Oxford University Press. https://global.oup.com/academic/product/kants-metaphysics-of-morals-9780198250104?cc=ca&lang=en&

Wood, A. W. (2008). *Kantian ethics*. Cambridge University Press. https://doi.org/10.1017/CBO9780511809651

Cite this article: Lee, B. D. (2026). The Categorical Imperative Against Lying and Its Permissible Violations. *Dialogue* 1–22. https://doi.org/10.1017/S0012217326101012

https://doi.org/10.1017/S0012217326101012 Published online by Cambridge University Press