

Original Article

**Kant and the King:**

**Lying Promises, Conventional Implicature, and Hypocrisy**

Roy Sorensen

Department of Philosophy, University of Texas at Austin

2210 Speedway, Stop C3500

University of Texas at Austin, Austin TX 78712-1737

and

Arché Research Centre at the University of St Andrews in Scotland

and

Ian Proops

Department of Philosophy, University of Texas at Austin

2210 Speedway, Stop C3500

University of Texas at Austin, Austin TX 78712-1737

Correspondence

Roy Sorensen

Department of Philosophy, University of Texas at Austin

2210 Speedway, Stop C3500

University of Texas at Austin, Austin TX 78712-1737

Email: [roy.sorensen@austin.utexas.edu](mailto:roy.sorensen@austin.utexas.edu)

### Abstract

Immanuel Kant promised, 'as Your Majesty's loyal subject', to abstain from all public lectures about religion. All past commentators agree this phrase permitted Kant to return to the topic after the King died. But it is not part of the 'at-issue content'. Consequently, 'as Your Majesty's loyal subject' is no more an escape clause than the corresponding phrase in 'I guarantee, as your devoted fan, that these guitar strings will not break'. Just as the guarantee stands regardless of whether the guarantor ceases to be your devoted fan, the compliance conditions of Kant's promise are not affected by Kant's ceasing to be the king's loyal subject. For good or ill, Kant made a lying promise to King Friedrich Wilhelm II in 1794.

### KEYWORDS

Grice, conventional implicature, hypocrisy, lying promise, mental reservation, posthumous harm

### 1 INTRODUCTION

The Enlightenment dissident, Immanuel Kant, evaded the theological censor in Berlin by submitting his work Religion within the Bounds of Bare Reason (1793) to the philosophy faculty at Jena, where it was duly approved (Laursen, 2015). Since part of the book had already been banned by the Berlin censors (Kuehn, 2001, 364-5), this amounted to an act of bad faith, and it was eventually treated as such by King Friedrich Wilhelm II and his ministers. Insulted, the King threatened through his minister Wöllner, to punish Kant for any further writing on religion (Letter to Kant of October 1, 1794). Kant disingenuously protested that his book made 'no appraisal of Christianity' (Pasternack, 2015). But Kant was ultimately contrite:

I find that, as Your Majesty's loyal subject [*als Ew. Majestät treuer Unterthan*], in order not to fall under suspicion, it will be the surest course for me to abstain entirely from all public lectures on religious topics, whether on natural or revealed religion, and not only

from lectures but also from publications. I hereby promise this. (Draft of a letter of October 12, 1794; Kant 1900, 11: 527-530)

The German—very literally translated—is: ‘I hereby thereto bind myself’ (Ich ... mich hiermit dazu verbinde), and ‘thereto’ refers to the aforementioned abstaining. Kant is binding himself to a course of action. He is not merely warranting the truth of a sentence.

Kant’s friends were relieved that he compromised on his commitment to scholarly free speech. They thought his forthright and indeed explicit promise was needed to save his position at the university. The King and his counter-Enlightenment minister would be able to see through any circumlocution. The only way to lift the threat was clear and complete capitulation.

As Kant had been hoping, the King died before him. Kant lost no time in publishing the manuscript waiting in his desk, along with other essays on religion. In the *Conflict of the Faculties* (1798), he tries to pre-empt any suspicion that his absolute opposition to insincere promises was hypocritical. Publishing a slightly different draft of the letter we quoted from above but with the same general gist -- and the same promise -- Kant proudly portrayed himself as having, through his careful choice of wording, cleverly avoided a lifetime commitment to refraining from publishing on religion (7: 7-10). We are grateful beneficiaries of Kant’s determination to keep publishing, and no less of his resistance, through his general defense of the freedom ‘of the pen’, to a dangerous opponent of the Enlightenment. Nevertheless, we contend that, on the assumption that the letter Kant sent matches the draft from which we quoted, he is guilty as charged.

Our purpose in making this contention is not to engage in idle gossip or to score petty points against a great thinker. It is rather to construct a novel argument against Kant’s extreme

stand on the impermissibility of lying. We will argue that Kant in fact lies with no more compunction than the ambitious, busy academics of his era. This lack of inhibition, though moderate in degree, is sufficient to constitute evidence that he does not in fact believe his theory concerning the absolute wrongness of lying. Kant's actual beliefs about lying align with those of his friends and friendly colleagues (and help to explain how he manages to have friends and friendly colleagues). This difference between belief and professed belief matters; if an expert ethicist does not believe his own ethical theory, this counts as evidence against it. These points will require careful argument and we do not expect them to be obvious at this stage in our discussion. Nonetheless, they should, we hope, help to explain why we are bothering to argue that Kant lied in the present case and in others.

We turn now to substantiating our charge that Kant lied to the king. Our first task is to defend the coherence of this charge.

## 2 IS 'LYING PROMISE' A DOUBLE OXYMORON?

We are claiming that Kant made a lying promise to the king. 'Lying promise' is the head-scratching English translation of Kant's 'Lügenhaftes Versprechen' and also of 'unwahres Versprechen' (see *Groundwork* (1785) at 4: 403). The latter phrase may be literally translated as 'untrue promise', but modern Germans are apt to hear it as having a meaning we'd render in English as 'false promise'. Such a locution raises the suspicion that false promises are not promises (just as false friends are not friends). The suspicion is reinforced by the following train of reasoning: A pretend promise is not a promise. Consequently, sincerity is a necessary condition for a making a promise. Therefore, no promise is made when the promisor is untrue to his own belief about whether he will keep his promise. In a false promise, continues this line of

reasoning, the speech act of promising misfires. Thus, one might conclude that 'lying promise' is an oxymoron on the grounds that no lying promise is a *promise*.

Alternatively, one might seek to argue that no lying promise is a *lie* (Meibauer 2014, 76). For there is a ready syllogism: *All lies are assertions. No promise is an assertion. Therefore, no promise is a lie.* According to James Mahon (2010, p. 50), all twentieth-century definitions of 'lie', since they presuppose that lies are without exception assertions, actually preclude lying promises -- often to the surprise of the definers. Fortunately, Kant is innocent of this charge, for he requires only that a lie be a *declaration* and he does not require that a declaration be an assertion. A declaration, in Kant's use of that term, can be either an assertion or a promise (MM 6: 429).

Attempts to exclude 'lying promises' from the class of promises may date to ancient Greece. In the *Theaetetus* (187-200), Plato portrays Protagoras as modeling false judgements on false eyes. The 'eyes' on the back of a pygmy-owl's head differ from the true eyes on its front. Sophists could extrapolate in a legally interesting direction: The true (that is, genuine) promisor believes he will do as he says. By contrast, one who promises what is not, being a 'false promisor', is not promising!

Plato soundly refutes Protagoras by localizing the 'false' in 'false belief' to the falsehood of the proposition believed. A false belief is just a belief directed to a proposition that has the truth-value: false. In 'false promise' 'false' is not functioning as a debunking adjective as it is in 'false scorpion'. To falsely promise is just to promise what one expects to be a false proposition. Although the pretend promises Ronald Reagan made while acting on stage were not promises, the insincere promises he made while serving as President were promises.

J. L. Austin agrees with Plato's insight about the function of 'false' in 'false promise'. And he also agrees that false promises are promises. Nevertheless, Austin still denies that insincere promises are lies or misstatements. Whereas *constative* statements, such as 'The door is closed', fit the philosophers' stereotype of meaningful language as reporting facts, *performative* statements, such as 'I order you to close the door!' and 'I promise to close the door', are actions rather than reports. Performatives challenge the contrast between words and deeds. Austin claimed that performative statements such as Kant's promise to the King lack truth-values. The promisor purports to make the world fit the sentence by bringing a promise into existence. The predictor purports to make the sentence fit the world by fitting his words to what is foreseen. Promising to do something you do not intend to do is 'perhaps misleading, probably deceitful and doubtless wrong, but it is not a lie or a misstatement' (Austin, 1962, p. 256).

Austin eventually conceded that assertion passes the 'hereby' test for performatives and accordingly abandoned the performative/constative distinction (1962, p. 150). Grice (1989, p. 122 and p. 362) stresses that multiple speech acts can be performed with a single utterance. If an explicit promise is also an implicit prediction, then a lying promise is a lie by virtue of nonbelief in the prediction. On these same assumptions, 'I promise to not drink but I predict I will drink' is inconsistent. The explicit promise I will not drink, insofar as it is at the same time an implicit prediction that I will not drink, contradicts the explicit prediction I will drink.

The promisor is in a privileged position to make the prediction based on his intention to fulfill the promise. By announcing the promise, the speaker clarifies the nature of the assertion he is making – one based on his own resolution to make the assertion true. This explains what is insulting in the following exchange: Promisor: *I promise not to drink*. Tactless friend: *Good, but will you drink?*

The promisor might try to climb over the insult: 'I promise not to drink despite predicting I will drink. The first step in overcoming an addiction is acknowledging loss of control.' But this incontinent compatibilism is no less bewildering than a soldier promising his comrades that he will not break under torture while simultaneously predicting he will break. A soldier can *resolve* to not break without predicting he will succeed. But a promise to comrades is an assurance; the soldier is inviting them to make plans around the fulfillment of his promise not to break. Resolution is weaker than prediction. Prediction is weaker than promising. A soldier may be in a position to predict that she will be in the barracks without being in a position to promise she will be in the barracks.

A logical explanation for the superior strength of a promise is that it entails a prediction. For instance, Neri Marsili (2016, p. 285) proposes a model of illocutionary entailment that systematically deduces assertions from promises. He notes that this model vindicates folk intuitions. Almost all research subjects agree that Coco lies to Baba if Coco promises to not drink while neither believing nor intending that she will not drink (Marsili, 2016, section 4). Far from being double oxymorons, lying promises are both lies and promises. The lying promisor is in double trouble.

Those less attached to 'All lies are assertions' could allow other speech acts to be the bearers of 'lie'. 'Assertion' is a term of art in Anglophone philosophy. Speakers of other languages than English may see less profit in maintaining assertion's monopoly on lying.

Conveniently for our argument, Kant is in this permissive group. According to Kant, promises are declarations even if they are not assertions and lies are a species of declaration. For our purposes what matters is that Kant is committed to: 'All lying promises are lies'. He would not welcome a rescue that proceeds: *Contrary to Kant's moral philosophy, 'lying promise' is a*

*contradiction in terms. Given the impossibility of telling a lying promise to anyone, Kant did not lie to the King (although Kant may have made a promise that he did not intend to keep).*

### 3 IS 'COERCED PROMISE' AN OXYMORON?

Plausibly Kant's promise to the king was coerced. Some deny that a coerced promise is a promise. But Kant maintains in his lectures on Natural Law of 1784 that a coerced promise is strengthened by an oath and ought (as a matter of ethical -- but not juridical -- duty) to be kept.<sup>1</sup> Thus he says:

If one has been coerced into making a promise strengthened by an oath is he is obliged to keep it? Ethically yes, but not juridically for the promisor has no right to demand it of me because he can acquire nothing **in injusta** [unjustly] (27: 1366).<sup>2</sup>

For Kant, although the coercion is unethical, the coerced promise still creates an ethical obligation on the promisor -- or at least that is the case so long as what is promised is itself morally permissible (as not publishing on certain religious questions would seem to be).<sup>3</sup> A wrong promise is still a promise even if a wrong lie is not really a lie.

---

<sup>1</sup> Kant's immediately subsequent discussion makes clear that the strengthening provided by the oath is not a strengthening of the obligation created by the promise but rather only of one's consciousness of that obligation (27: 1366).

<sup>2</sup> Kant says that no *juridical* obligation is created. For he denies that the coerced promise creates a *special* right of a particular person. Consequently, no particular person acquires a claim against me in virtue of my coerced promise (just as no one acquires a legal right of ownership upon receiving stolen goods). This detail, however, need not detain us. Kant is clear that an *ethical* obligation *is* created.

<sup>3</sup> The point of this qualification is worth explaining. Suppose I am coerced into promising to tell a lie. If I must—as a matter of ethical obligation—keep my promise, then absurdity results: I am ethically obliged to act unethically (at least by Kant's lights). More generally, even a voluntary promise to act unethically ought not to generate an obligation. And yet Kant seems to hold the reverse. For if the coercion fails to release one from the obligation, it is hard to see how its absence could bring about that release. Noticing this problem with promising what is contrary to duty, Ryan S. Kemp recommends that we suppose that a promise to



Kant claims to have outwitted the King by craftily inserting the clause ‘as your Majesty’s loyal subject’. ‘This phrase’, he confesses, ‘was chosen by me most carefully, so that I would not be deprived of my freedom ... forever, but only so long as His Majesty was alive’. (Note to the account given in *The Conflict of the Faculties*, 7:10).

This hoped-for loophole is, however, a mirage. The non-clausal appositive, ‘as your Majesty’s loyal subject’, is not part of the ‘at-issue content’ of the primary assertion.<sup>4</sup> The material in the phrase is therefore irrelevant to the compliance conditions of the promise.

The material in *clausal* appositives does affect at-issue content: ‘Not only do I acknowledge you as my sovereign, but I also promise to abstain from further public lectures on religious topics’. The *p* in the ‘Not only *p* but also *q*’ construction is backgrounded. Yet this same *p* does nonetheless function as at-issue content – as is manifest from the clause hosting the performative verb ‘acknowledge’ (Horn, 2013, pp. 164-165). Clausal appositives represent secondary speech acts.

Non-clausal appositives, by contrast, are Fregean asides that make no difference to the sense or reference of the primary assertion. They facilitate communication. They add tone, scent, or color. But they are impotent as escape clauses. Illustration: the pay-out conditions for ‘I bet you, as a loyal German, that Germany will win more gold medals than France in 2024’ are not

---

perform an action which the promisor knows to be immoral is not after all a promise (2019, pp. 462-3). Perhaps this is a reasonable thing to say on Kant’s behalf. But Kemp offers no texts that suggest that Kant actually adopted such a line – and we know of no such evidence. Kemp also suggests that Kant might permit a promise to act immorally to be broken. Kemp’s arguments on this score are textually substantiated and seem to us more persuasive – hence our qualification ‘or does so as long as what is promised is itself morally permissible’.

<sup>4</sup> Horn aligns with Gottlob Frege’s (1892, p. 38) discussion of Napoleon in ‘Über Sinn und Bedeutung’ in opposition to Christopher Potts’s (2005) radical overhaul of conventional implicature. The taxonomy of not-at-issue contents is controversial. Thorsten Sander (2022) even denies that there is any unique and general taxonomy.

affected by whether the speaker is a loyal German.<sup>5</sup> The bettor is committed to the wager even if he discovers himself to be French! The point applies to all commissive speech acts: vowing, opposing, offering, and so on.

People often try to beef up the credibility of promises by supplying preambles, footnotes, marginalia, and addenda. But the truth-values and sincerity of these supplements are irrelevant to the fulfillment conditions of the promises in question. A closet atheist who says ‘As God is my witness, I promise to repay the loan’ has made a promise. If he plans not to pay, he has made a lying promise even if he fancies himself a clever evader of the debt.

H. P. Grice (1989) repackages many of Gottlob Frege’s insights in his theory of conventional implicature. But scholars who deny that Kant made a lying promise, from the gentle Alasdair MacIntyre (1995) to the stern Bernard Williams (2002), have instead applied Grice’s famous theory of *conversational implicature* thus: *Kant’s promise that he will stop publishing on religion carries the conversational implicature that he will stop forever. But this implicature is not part of what was said (in Grice’s favored sense). Therefore, Kant did not break his promise when he resumed publishing after the King died. The distinction between lying and misleading is of crucial moral importance for Kant. Consequently, Kant’s ruse succeeded in protecting his intellectual freedom without violating his universal prohibition against lying promises.*

The italicized paragraph vindicates the common-sense distinction between lying and telling misleading truths.<sup>6</sup> This is a persuasive rejoinder to J. H. Fichte’s definition of lying as

---

<sup>5</sup> In the first *Critique*, Kant pioneers the use of bets as measures of belief (A824/B852). Some psychologists explain heightened logical performance on contract reasoning by postulating a cheater detector module. Economists use rational choice theory to explain why predictions are cheap talk but bets are serious discourse.

<sup>6</sup> Jennifer Saul’s Lying, misleading, and what is said continues the Gricean vindication at book length. Saul notes that a secure distinction between lying and misleading is only the first step in defending the common

any intentional deception. Fichte contends we all have a duty to perfect one another.<sup>7</sup> Since errors impair freedom, we have a duty of communication (Bacin 2021). Any breach of this duty is a lie—even telling a misleading truth (1798, p. 270). By Fichte’s lights non-verbal actions that are intended to mislead are also lies. Indeed, one lies by omission when one refrains from correcting an error that one did not cause. All of these lies are absolutely forbidden (Nease 2002).

When compared with Fichte, Kant’s ‘never-lie’ purism begins to seem relatively moderate. Whereas Kant merely denies that one may lie in certain cases where lying is in fact intuitively permissible (for example, when this seems necessary to save an innocent man from murder), Fichte goes so far as to castigate even those who are merely tempted to lie: ‘The true root of your perversion’, he scolds, ‘lies in the fact that such a lie *even occurred* to you as a possible way of escaping certain predicaments and that you are now able seriously to consider whether one might be allowed to avail oneself of such a means of escape’ (1798, p. 274). Again, whereas Kant is prepared to explore the possibility that some lying might be necessary, Fichte firmly condemns any such inquiry: ‘The possibility of such an escape route does not even occur to an honest person, nor [if ethicists were honest people] would an investigation into the morality of the white lie have entered into ethics.’ (1798, p. 275). In view of this strictness, it’s clear that Fichte would have reached the conclusion that Kant lied to King Friedrich Wilhelm II. We believe he would have been right, albeit for the wrong reasons.

---

sense assumption that the distinction is *morally* relevant. Indeed, Saul shocks most readers by arguing against the *moral* relevance of the distinction.

<sup>7</sup> In the Metaphysics of Morals, Kant disagrees, ‘it is a contradiction for me to make another’s *perfection* my end and consider myself under obligation to promote this. For the *perfection* of another human being, as a person, consists just in this: that *he himself* is able to set his end in accordance with his own concepts of duty; and it is self-contradictory to require that I do (make it my duty to do) something that only the other himself can do.’ (6: 386) An anonymous referee alerted us to Fichte - and to subsequent subtleties about ‘saying as’.

With the benefit of hindsight, we can see that Gricean considerations suffice to refute Fichte's doctrine that intentionally deceiving with misleading truths suffices for lying. And yet, in spite of its success, the Gricean refutation of Fichte fails to save Kant. Non-clausal appositives carry *conventional*, rather than conversational, implicatures. As Grice (1989, p. 75) explains in a brief aside of 'Logic and Conversation', *conventional* implicatures are semantic rather than pragmatic. They leave the truth conditions of the primary assertion unaffected. In this respect, conventional implicatures resemble footnotes. Grice thinks this explains the difference between 'but' and 'and'. 'She is poor but honest' commits the speaker to a contrast despite having the same truth conditions as 'She is poor and honest'. Unlike conversational implicatures, this conventional implicature cannot be cancelled by saying 'She is poor but honest which is not to say that her being honest is in any way surprising given her poverty'. Grice's other illustrations are also single words: *even, although, moreover, therefore*, plus the phrase *on the other hand*. Other philosophers add: *managed, bothered, bastard, goodbye*, slurs, indicative conditionals, the active/passive distinction, and finally, formal/informal address as in the French *vous/tu*. The linguist Christopher Potts thinks complex expressions yield the best specimens: 'Supplements (appositives, parentheticals) are the finest advertisement' for the existence of conventional implicatures (2005, p. 13).

The word 'this' in 'I promise this' refers to the content of the promise, not the marginalia about Kant being a loyal subject and his interest in avoiding the King's suspicion. From a logician's point of view, this side-talk is illuminatingly tuned out, yielding: ' . . . to abstain entirely from all public lectures on religious topics, whether on natural or revealed religion, and not only from lectures but also from publications. I hereby promise this.' That's what the King heard – and what Kant hoped the King would hear.

Material in a non-clausal appositive can be relevant to whether the promisee should *release* the promisor from the under-researched promise. Mistaken not-at-issue content may also be relevant to an *excuse* for breaking the promise. Perhaps the clause is even relevant to a *justification* for violating the promise. But not-at-issue content is not relevant to whether the promise is broken. Kant made a lying promise even if he imagined he had cunningly achieved sincerity.

#### 4 EXPIRED PROMISES

Kant denies that a promise expires with its promisee. In the course of making the more general point that moral knowledge is easy to come by, he invites the reader to pose the following hypothetical question to a boy of eight or nine years old (8: 286). A poor father promises to return money that a rich man has left for safekeeping. The rich man unexpectedly dies. None of his heirs needs the money. If the heirs received it, they would waste it. No one will notice if the poor father spends the deposited money on his family's urgent needs. May the father keep the money? Kant thinks it plain that the boy will answer 'no'.

On cases like this most philosophers agree with Kant's boy. They believe that promises survive the death of the promisee so long as they still remain executable at that point (Taylor 2021, 81-84). A favorite example of so-called 'posthumous harm' is a broken promise about matters beyond the lifetime of the promisee.<sup>8</sup> You promise your sole companion on a desert island to tend his flowers after he dies (Narveson, 1963, p. 63). He dies. You are obliged to keep the promise even if you know no one will learn you did not.

---

<sup>8</sup> Kant thinks dead people are people, hence ends in themselves. This entails respect for their privacy and reputation. Falsely accusing a dead neighbor wrongs *the neighbor*. (When? Kant hints; while the neighbor was alive because his interests extend to post mortem states of affairs.) Even true gossip is disrespectful because the dead are being used as a means of entertainment.

The utilitarian Henry Sidgwick (1874 section 3.6 sec. 9) claims that consensus for 'Promises must be kept' wanes for deathbed promises. He suggests that promises are binding only as long as the promisee can release the promisor. Sidgwick would advise Kant to ensure that the King is the only promisee.

## 5 DELIBERATELY AMBIGUOUS PROMISES

Kant does seem to treat his promise to the King as if he had said 'I hereby promise this [to you]' with the elided content, 'to you', intended to be understood by the addressee from the context. The promise would then, under Sidgwick's smartening up of current practice, lapse with the death of the receiver of the promise.

However, Kant intended the King to read the lifetime guarantee as applying to *Kant's* life-time rather than the King's. In contract law, ambiguities are interpreted against the drafter (Solan, 1993, pp. 87-92). The speaker's advantage in crafting the agreement is balanced by the hearer's right to resolve ambiguities. (This explains why your adversary's lawyer urges that you attest to having co-authored the contract.) Consequently, a drafter cannot override the hearer's preferred reading by adding further readings. The hearer is king, in this case: King Friedrich Wilhelm II.

To instead give precedence to the drafter of a contract, Kant could try the doctrine of 'mental reservation'. According to this private semantics, the content of the assertion tracks inner speech rather than the outer speech accessible to the audience. When the English-Reformation 'priest hunter' asks the recusant 'Are you a priest?', the Catholic priest can answer 'I am not a priest [of Apollo]'. The silently completed sentence is true and so not a lie to the only other being who knows what was said – God. When Kant was working out his views on lying

promises, the Jesuits were in the process of extending this doctrine about lying to promises. Roussel de la Tour (1763) effectively suggests that one who seeks to wriggle out of a payment should put their promise to pay in the form 'I promise to pay [to you or someone else]', where the dative phrase indicates to whom the promise is addressed. The sum expected by the addressee of this promise could thus be directed to a confederate. Pressed with the query 'Have you promised?', you can answer 'I have promised [though not with a promise that obligates me]'

Many Catholics disapproved of the doctrine of mental reservation, most famously, Blaise Pascal:

Listen, then, to the general rule laid down by Escobar: Promises are not binding, when the person in making them had no intention to bind himself ... so that when one simply says, I will do it, he means that he will do it if he does not change his mind; for he does not wish, by saying that, to deprive himself of his liberty. (Pascal 1656: p. 101; see also Filiutius 1633 in Roussel de la Tour 1763: p. 419)

If Kant's plan was to be Jesuitical, he would have executed it competently (and refrained from condemning silent mental reservations in the opening of 'Perpetual Peace'). His written promise was not scaffolded for application of the doctrine (which requires a sentence that can be lengthened by an unspoken escape clause). Kant had years to prepare for his turn at royal persecution. In other drafts of the letter, he experiments with false *conversational* implicatures. Instead of 'I hereby promise this', he merely expresses the belief that the surest way to obviate

the suspicion is for him to declare solemnly that etc. (7: 10). Kant probably foresaw that the King's ministers would have then pressed for a clarification.

When the promisee is no longer around to police compliance, the promisor's memory is free to reconstruct the promise. Unfortunately for Kant, the King got the promise in writing (though the ultimate letter may have been destroyed during World War II). All Kant could do was talk over what was silently but still legibly on record.

The over-talking did manage to persuade all those scholars who have previously volunteered a verdict. Alastair MacIntyre (1995, p. 337) summarizes, 'Kant succeeded in misleading the Prussian censors without lying, something he thought it morally important to do.' Kant's sternest critic, Bernard Williams (2002, p. 107), merely adds a stinging caveat: Kant 'fetishizes assertion' by adopting a conception of truthfulness that subtracts conversational implicatures. Williams only disagrees with MacIntyre and Adler (1997, p. 144) about the moral significance of the lying/misleading distinction.

Thomas Carson (2010, p. 78) contends that Kant trivializes his opposition to lying with Jesuitical applications of the doctrines of equivocation and mental reservation. But Kant is not speaking in his own voice when he lectures on Catholic word magic. He is adopting a perspective he has not bothered to seriously study. Consider Kant's friendly word for equivocation in the Collins lectures:

Aequivocatio is permitted, in order to reduce the other to silence and get rid of him, so that he shall no longer try to extract the truth from us, once he sees that we cannot give it to him, and do not wish to tell him a lie. If the other is wise, he will also let it go at that. It is quite difficult, though, to employ equivocation when we state and declare we are



expressing our views, for in that case the other may infer something from the equivocation, and then I have deceived him. (27: 449)

Kant is equivocating on 'equivocal'. Instead of talking about ambiguity he is really talking about generality.<sup>9</sup> If pumped for information, you may emit evasive generalities to signal that no details will emerge (by flouting Grice's submaxims of manner 'Avoid obscurity of expression' (for generality) and 'Avoid ambiguity' (for equivocation)). The nosy inquirer gets the message and relents. Success: You avoid both lying and silence. But Kant's recommendation to speak in generalities is not an endorsement of the doctrine of equivocation.

During emotional storms, ethicists do lapse into rationalizations that they earlier dismissed in the quiet of their studies. These performance errors do not create sincere speakers. Scolded by the King and fearing His Majesty's threat of 'unpleasant measures', Kant was desperate to de-escalate. He explored many possible exits including some precluded by his theoretical texts.

## 6 SAYING AS

In 'What is Enlightenment?', Kant distinguishes between what one says *as a clergyman* from what one says *as a scholar*. A pastor

---

<sup>9</sup> Ambiguity is the possession of multiple meanings. Generality is a matter of there being many states of affairs that could make the sentence true. For instance, when a woman responds to the cross-examiner's 'How old are you?' with 'I am over 21 years old', she commits to being an adult for legal purposes but frustrates his curiosity about which of many ages makes her answer true. Arnold Zwicky and Jerrold Sadock (1975) review the scope and limits of diagnostic tests for ambiguity.

. . . is bound to deliver his discourse to the pupils in his catechism class and to his congregation in accordance with the creed of the church he serves, for he was employed by it on that condition. But as a scholar he has complete freedom and is even called upon to communicate to the public all his carefully examined and well-intentioned thoughts about what is erroneous in that creed and his suggestions for a better arrangement of the religious and ecclesiastical body. And there is nothing in this that could be laid as a burden on his conscience. For what he teaches in consequence of his office as carrying out the business of the church, he represents as something with respect to which he does not have free power to teach as he thinks best, but which he is appointed to deliver as prescribed and in the name of another. He will say: Our church teaches this or that; here are the arguments it uses. He then extracts all practical uses for his congregation from precepts to which he would not himself subscribe with full conviction but which he can nevertheless undertake to deliver because it is still not altogether impossible that truth may lie concealed in them, and in any case there is at least nothing contradictory to inner religion present in them. (8:38)

This passage suggests a loophole. Kant promised *as Your Majesty's loyal subject* to cease publication on religion but not *as a scholar*. He may promise to stop disseminating his views *as a teacher of the youth* (which for Kant is in the private sphere – and is therefore subject to censorship) while still disseminating them *as a thinker in the republic of letters* (which is in the public sphere – and therefore ought to be free of censorship). This rescue generalizes. Kant promises *as rector* of the University of Königsberg to attend church on ceremonial occasions.

But when Kant repeatedly claims to be indisposed (Bird 2006, p. 26), he is not repeatedly breaking his promise *as an independent thinker*

By means of `as'-clauses we sometimes try to avoid contradictions. Consider a college professor who is asked by a colleague with a broken leg to look after her cat while she recovers. Concerned that his compliance might inflate expectations for collegial behavior, and fearing for his workload, he negotiates a balanced response: `I do not promise to mind your cat as your colleague, but I do promise to do so as your friend'. Arguably, the material in the `as'-clause becomes at-issue content.

Nevertheless, Kant constrains the contradiction-dissolving power of `as' clauses. For he denies that the pastor can preach a doctrine that he believes to be definitely false. This limit also conflicts with Kant's assurance that a sermon merely reports church doctrine. When an anthropologist reports the doctrine of the trinity, he does not commit to the possibility of the doctrine being correct. But Kant concedes that the pastor in the pulpit does so commit.

## 7 HYPOCRISY AND THE RELEVANCE OF BIOGRAPHY

After the King died, the clouds that had been hanging over Kant dispersed. But now a new worry threatened to darken the newly sunlit uplands: would resuming publication lead him to be scorned as two-faced?

The moralist is a critic who looks down on a wrong-doer and so must stand in judgment above him. A hypocrite violates the norm that the blamer must have standing to blame. He loses standing by being guilty of the very offense he criticizes. According to Patrick Todd, `one has moral standing to blame if and only if one is morally committed to the values that condemn the wrongdoer's actions.' (2019, p. 357). Commitment is `endorsement of the relevant value as

value, and at least some degree of motivation to comply with it.' (2019, p. 357). Endorsing honesty generates testable consequences.

Hypocrisy is only one way criticism can fail to be appropriate. Criticizing without sufficient evidence is inappropriate but not thereby hypocritical. *Expressing* criticism of matters that are not one's business is inappropriate but not thereby hypocritical. At times, Kant worried that his moral criticisms might be hasty or meddling. But in the case of the contemplated resumption of public criticism, his worry was hypocrisy.

In the *Critique of Practical Reason* Kant had seconded the ancients' intolerance of hypocrisy. They believed that no man could call himself a philosopher, a lover of wisdom, 'unless he could establish [philosophy's] infallible effect on his own person as example' (5: 109, Beck translation). Kant would not explain systematic departures from one's own precepts as weakness of will: 'What I recognize immediately as a law for me', he says, 'I recognize with respect.' (1959, pp. 20-21). And to respect the law is, among other things, to be motivated by it. After all, respect is 'immediate determination of the will by means of the law and consciousness of this [immediate determination]' (G 4: 401 note †).

On these assumptions, an uninhibited liar cannot believe that all lies are forbidden. And so, when we discover his lack of inhibition, we deny he believes his absolute prohibition. If an ethicist does not believe his ethics, moreover, then any weight assigned to his expertise counts against his ethics rather than in its favor.

The evidential value of hypocrisy is plausible to all moral internalists. To count a proposition as a moral reason, one must be motivated to act on it.

Moral *externalists*, by contrast, think the uninhibited liar could as easily believe 'All lies are forbidden' as 'All squares are equilateral'. According to the moral externalist, the discovery

that Kant lied to everybody, from cobblers to kings, would not diminish the plausibility of his absolute prohibition of all lies as grave misdeeds. Kant could still know all lies are forbidden and criticize others on this basis. Objecting ‘But you yourself lie without compunction!’ would be an instance of the *tu quoque* fallacy. Only a fool refuses to learn from hypocrites. Only a fool gives extra credence to a morality on the grounds that the moralist practices what he preaches—or so says the externalist.

Kant, by contrast, did think that biography is relevant – and certainly thought it would be regarded as relevant by potential readers. The King may have regarded Kant’s promise as insurance regardless of whether it was kept. The King may have reasoned: Either Kant breaks his promise (and ceases to corrupt the King’s subjects because he is now recognized as a hypocrite) or Kant keeps his promise (and ceases to corrupt the King’s subjects because he keeps silent). Either way, the King’s subjects are protected.

To prevent publication from being self-defeating, Kant needed to persuade readers that he was not thereby revealing that he had made a lying promise to the King. The year following the King’s death (the King died in mid-November 1797), Kant published an account of the incident in the Preface to his compilation of previous work, *The Conflict of the Faculties*. There, wanting to portray himself as a victim of his own candor, he explained that he had been reproached by the King simply because, having nothing to hide and ‘not wanting to be accused of using underhanded means’, he had followed his usual policy of signing his name to ‘all my writings [*Schriften*]’ (7: 6). Admirably forthright! Or so it *seems*; for the claim that he had such a policy turns out to be far from true. Only eight years earlier, he had indeed failed to ‘sign his name’ to an anonymous work, namely: ‘On the propensity to fanaticism and the means to oppose it’ (1790). It is unlikely that Kant could, in so short a time, have forgotten that he took this

precaution—unlikelier still that he could have forgotten each one of the *nine* previous occasions on which he had taken this same precaution. (For a list, Steve Naragon's 'Kant in the Classroom' website: <https://users.manchester.edu/facstaff/ssnaragon/kant/Helps/KantsWritings.htm>). Nor are these the only instances in which Kant lied. Another central example is his claim to have made only cosmetic revisions 'in the manner of exposition' in the second edition of the first *Critique*. Kant must have known that several of his revisions were substantive (Proops and Sorensen, 2023).

We agree with moral externalists that hypocrisy can be overrated as evidence of non-belief. It would be absurd to take an occasional lie by Kant as decisive evidence that he did not really believe his theory. And no less absurd to further conclude that the theory was therefore false -- since disbelieved by an acknowledged expert on matters ethical. Yet we agree with Kant and his fellow internalists that hypocrisy is still *some* evidence of insincerity.

Kant cautions that those who believe that there are necessary lies should not admit this to children: 'they would soon take the smallest excuse for this necessity, and often allow themselves to tell lies' (VP 9: 940). If a mother goes further and declares that all lies are wrong, then she is not engaged in rhetorical exaggeration. Hyperbole requires the speaker to expect the audience to recognize the overstatement (specifically by *flouting* Grice's (1989, p. 27) maxim of quality, '1. Do not say what you believe is false. 2. Do not say that for which you lack evidence.'). When her children grow up and notice their mother telling lies, she might replace 'Never lie' with a weaker prohibition. Alternatively, she might keep up appearances or even become a more strident denouncer of lies. As an eighteenth century professor, Kant had pastoral duties toward adolescent boys. His audience never grew up. Kant had accumulated a half

century of precedent to continue strong prohibitions against lying, theft, and masturbation (worse than suicide because suicide at least requires courage).

Allen Wood remarks, 'If one has a taste for malicious irony, it is possible to present this situation in an especially unflattering light if we say that as a moralist, Kant permitted himself to make statements (and on the morality of truthfulness, of all subjects!) that were *knowingly false*.' (2007, p. 330 n. 27). This malice is less appetizing when one reflects on the pressure on one's own parents to make morality practical to children. When compared to Fichte, Kant emerges as a moderate who ingeniously rescues politeness and dinner parties from the choking hazard of 'Never lie'.

Nevertheless, if Kant was as hypocritical as his first biographer intimates, then one could conclude that Kant's ethics fails by its own standards.<sup>10</sup> A morality that endorses the relevance of hypocrisy makes hypocrisy relevant for its own case. So even a moral externalist must concede that failure to practice what one preaches can decrease the plausibility of what is preached.

Kant puts himself in a difficult position. His condemnation of lying is extreme in its content (a universal generalization), its expressive intensity, and its persistence. Yet, his behavior manifests no more aversion to lying than the baseline set by other academics of his era. A biographer who is trying to ascertain whether Kant lied in a given case gets little guidance from Kant's absolute prohibition. Consider J. H. W. Stuckenberg who quotes from a newspaper in 1848 reporting that 'In [Kant's] last years, his conscience troubled him because at one time, in order to decline a disagreeable invitation, he pretended to be already invited for the time

---

<sup>10</sup> The first biographer was Kant's colleague and professor of medicine, Johann Daniel Metzger. Metzger portrays Kant as a great philosopher but only an average man. Average men are averse to lying but not to the extent of believing all lies are forbidden. The alleged gap between Kant's extraordinary pronouncements and his ordinary practice led to three biographies by former students who coordinated to save Kant's good name (Kuehn 2001, 7-8). These pastors stress Kant's clock-work consistency.

designated' (second edition, 1882, p. 446 n. 132). Stuckenberg cannot dismiss the report on the grounds that Kant's ethical theory precludes the lie. Did Kant lie to secure posts for his disciples (Kuehn, 2001, 316)? Did Kant suborn Christian Jacob Kraus to argue insincerely for critical philosophy (Kuehn, 2001, 331)? To answer such questions the biographer, rather than focusing on Kant's ethics, should focus on the fact that social and scholarly ambitions lower inhibitions about lying.

Kant's level of inhibition may have dipped to its lowest point for politics. Possibly, Kant did not care whether he made a lying promise to *a counter-Enlightenment monarch*. After all, Kant had earlier flattered a *pro-Enlightenment monarch* (Laursen 2015, 104). Frederick 'the Great', to whom Kant dedicated the *Critique of Pure Reason*, is told that strong rulers do not need censorship. Nor do they have an interest in becoming guardians of religion. And strong leaders have no fear of shadows cast by the arts and sciences (8: 41). Despite elsewhere favoring republics, Kant tells Frederick that his reign demonstrates how a free press is better protected by a monarch's fiat than a republican election. Kant panders to Frederick's contempt for religion. Kant is a political gradualist who hopes to nudge all monarchs onto a slippery slope. His ideas about censorship, book piracy, and education are intended to persuade royalty to begin a gradual descent. These enemies of maturity gently reach bottom before realizing they were taken for a ride. The last sentence of 'Perpetual Peace' assures the sovereign, 'And since the class of philosophers is by nature incapable of forming seditious factions or clubs, they cannot incur suspicion of disseminating propaganda.' (1991, p. 151)

It is also possible that Kant did feel reluctant to make a lying promise to Friedrich Wilhelm II – but not enough to steer clear of any risk. After the promise and before the King's



death, Kant continued to publish remarks on religion that bordered on breaking his promise (Kuehn 2001, p. 402). Sincere promisors maintain a margin of safety.

Either of these hypotheses is compatible with this essay's thesis. For good or ill, Kant made a promise to a menacing monarch that he did not intend to keep. As is familiar from contract law, you may make a mistake about what it was that you promised. Your intention to keep your non-existent weak promise is compatible with your intending not to keep the strong promise you actually made. Kant intended to return to the topic of religion if he survived the King. Fortunately for we admirers of Kant, this contingency plan was activated.

#### ACKNOWLEDGEMENTS

This essay has been improved by the comments and criticisms of Laurence Horn, James Mahon, Lawrence Pasternack, Jens Timmerman, Patrick Todd, and the audience of the Language and Mind Seminar at St. Andrews University on 16 May 2023. We have also benefitted from the most instructive referee reports we have ever received.

#### ORCID

0000-0003-1635-0980

#### REFERENCES

- Adler, J. (1997) . Lying, deceiving, or falsely implicating *Journal of Philosophy*, 94, 435–452.
- Austin, J. L. (1962) *How to do Things with Words*, Oxford: Oxford University Press 2<sup>nd</sup> edition 1975.

Bacin, S. (2021). My Duty and the Morality of Others: Lying, Truth, and the Good Example in Fichte's Normative Perfectionism. In Stefano Bacin & Owen Ware (eds.), *Fichte's System of Ethics: A Critical Guide*. Cambridge, UK: Cambridge University Press. pp. 201-220.

Bird, G. (2006). *A Companion to Kant* Maiden, Massachusetts: Blackwell.

Pascal, B. 1656. *Les lettres provincielles*. Manchester: The University Press, 1920. English translation by Thomas M'Crie in *Pensées [and] The Provincial Letters*. New York: Modern Library, 1941.

Carson, T. (2010) *Lying and Deception* New York: Oxford University Press.

Fichte, J. G. (1798). *The System of Ethics: According to the Principles of the Wissenschaftslehre*. Cambridge University Press, 2005

Frege, G. (1892). Über Sinn und Bedeutung, in *Zeitschrift für Philosophie und philosophische Kritik*, 100: 25–50.

Grice, H. P. (1989). *Studies in the Way of Words*, Cambridge, MA: Harvard University Press.

Horn, L. (2013). I love me some datives: Expressive meaning, free datives, and F-implicature. In Daniel Gutzmann & Hans-Martin Gärtner (eds.), *Beyond Expressives: Explorations in Use-Conditional Meaning*. Leiden: Brill. pp. 151-199.

Kant, I. (1900--) *Gesammelte Schriften* (Berlin: Königlich-Preussischen Akademie der Wissenschaften zu Berlin [now De Gruyter]). Cited by volume number and page.

\_\_\_\_\_ (1949) *Critique of Practical Reason and Other Writings in Moral Philosophy*, trans. Lewis White Beck, Chicago University Press.

----- (1785, 1969) *Foundations of the Metaphysics of Morals*, trans. Lewis White Beck, ed. Robert Paul Wolff, New York: MacMillan.

Kant, Immanuel (1991). *Toward Perpetual Peace*, in *Kant: Political Writings*, 2nd. ed., H Reiss, ed., Cambridge University Press.

Kant, Immanuel, (1996). *An Answer to the Question: What is Enlightenment?* reprinted in *Practical Philosophy*, translated by Mary Gregor, in *Immanuel Kant, Practical Philosophy*, Allen Wood and Mary Gregor (eds.), Cambridge: Cambridge University Press.

----- (1996). *Groundwork of the Metaphysics of Morals, The Metaphysics of Morals, and On a supposed right to lie from philanthropy*, Mary. Gregor (trans.), in Immanuel Kant, *Practical Philosophy*, Allen Wood and Mary Gregor (eds.), Cambridge: Cambridge University Press.

----- (1959). *Fundamental Principles of the Metaphysic of Ethics* tr. T. K. Abbot 10<sup>th</sup> ed. London.

----- (1997). *Lectures on Ethics*, P. Heath (trans.), Peter Heath and Jerome Schneewind (eds.), Cambridge: Cambridge University Press.

----- (2009) Kant, Immanuel (2009). *Religion Within the Bounds of Bare Reason*. Trans. Werner S. Pluhar, Indianapolis: Hackett Publishing Company.

Kemp, R. S. (2019). Kant on the bounds of promise making: A Mendelssohnian account. *European Journal of Philosophy* 27 (2):453-467.

Kuehn, M. (2001). *Kant: A Biography* Cambridge University Press.

Laursen, J. C. (2015). *Censorship from Rulers, Censorship from Book Piracy: The Strategies of Immanuel Kant. Censorship Moments: Reading Texts in the History of Censorship and Freedom of Expression*. Ed. Geoff Kemp. London: Bloomsbury Academic: 103–108.

MacIntyre, A. (1995). Truthfulness, Lies, and Moral Philosophers: What Can We Learn from Mill and Kant? in *The Tanner Lectures on Human Values*, 16: 307-361. Salt Lake City:

University of Utah Press

Mahon, J. E. (2010). The definition of lying, and why it matters. *Roundtable at the. School of Info Resources and Library Sciences.*

Marsili, N. (2016). Lying by Promising. A study on insincere illocutionary acts. *International Review of Pragmatics* 8 (2):271-313.

Meibauer, J. (2014) <sup>12</sup>When Jesus spoke again to the people, he said, "I am the light of the world. Whoever follows me will never walk in darkness, but will have the light of life."

<sup>13</sup>The Pharisees challenged him, "Here you are, appearing as your own witness; your testimony is not valid."

<sup>14</sup>Jesus answered, "Even if I testify on my own behalf, my testimony is valid, for I know where I came from and where I am going. But you have no idea where I come from or where I am going. <sup>15</sup>You judge by human standards; I pass judgment on no one. <sup>16</sup>But if I do judge, my decisions are true, because I am not alone. I stand with the Father, who sent me. <sup>17</sup>In your own Law it is written that the testimony of two witnesses is true. <sup>18</sup>I am one who testifies for myself; my other witness is the Father, who sent me."

) *Lying at the Semantics-Pragmatics Interface*. Berlin: Mouton de Gruyter.

Pasternack, L. (2015). Kant's 'Appraisal' of Christianity: Biblical Interpretation and the Pure Rational System of Religion. *Journal of the History of Philosophy* 53 (3):485-506.

Potts, Christopher (2005) *The Logic of Conventional Implicature*, Oxford University Press.

Proops, Ian & Sorensen, Roy (2023). Destigmatizing the exegetical attribution of lies: The case of Kant. *Pacific Philosophical Quarterly*. <https://doi.org/10.1111/papq.12442>.

Narveson, J. (1963). The Desert-Island Problem. *Analysis* 23 (3):63 - 67.

Nease, L. (2002) The Severity of the Moral Law in Fichte's *Science of Ethics* in *New Essays on Fichte's Later Jena Wissenschaftslehre* ed. Daniel Breazeale and Tom Rockmore Evanston, Illinois: Northwestern Press: 248-258.

Sander, T. (2022). Taxonomizing Non-at-Issue Contents. *Grazer Philosophische Studien* 99 (1):50-77.

Roussel de la Tour, C. P. (1763). Extraits des assertions dangereuses et pernicieuses en tout genre, que les soi-disans jesuites ont, dans tous les tems & perseveramment, soutenues, enseignées & publiées dans leurs livres. (Tome Second, Cap. XII: Parjure, Fausseté, Faux Témoignage.) Paris: Pierre-Guillaume Simon.

Saul, J. M.(2012). *Lying, misleading, and what is said: an exploration in philosophy of language and in ethics*. Oxford: Oxford University Press.

Sidgwick, H. (1874). *The Methods of Ethics* 7th ed. London: Macmillan, 1962.

Solan, L. M. (1993) *The Language of Judges*, University of Chicago Press.

Stuckenber, J. H. W. (1882) *The Life of Immanuel Kant* ([Macmillan 1982] Bristol: Thoemmes Antiquarian Book, 1990).

Taylor, J. (2021). Promises to the Dead. *Royal Institute of Philosophy Supplement*, 90, 81-103.

Todd, P. (2019). A Unified Account of the Moral Standing to Blame. *Noûs* 53:347-374.

Williams, B. (2002). *Truth and Truthfulness*, Princeton University Press.

Wood, A. W. (2007). *Kantian Ethics*, New York: Cambridge University Press.

Zwicky, A. M. and J. M. Sadock (1975). Ambiguity Tests and How to Fail Them in *Syntax and Semantics* (Volume IV), J. Kimball (ed.), New York: Academic Press, 1–36.

