

De-stigmatizing the Exegetical Attribution of Lies:

The case of Kant¹

By Ian Proops and Roy Sorensen

Abstract: Charitable interpreters of David Hume set aside his sprinkles of piety. Better to read him as lying than as clumsily inconsistent. We argue that the attribution of lies can pay dividends in historical scholarship no matter how strongly the theorist condemns lying. Accordingly, we show our approach works even with one of the strongest condemners of lying: Immanuel Kant. We argue that Kant lied in his scholarly work, and even in the first *Critique*. And we defend the claim that this lie attribution, strange as it may sound, amounts to a kind of scholarly charity.

Historians of philosophy are expected to extend ample charity to the philosophers they study—and rightly so. But in what exactly does this interpretive charity consist? We argue that, in addition to seeking to make the author come out as insightful, astute, reasonable, subtle, and consistent, the charitable interpreter should sometimes be prepared to attribute lies. We are aware that this will probably sound counterintuitive. There is a powerful taboo on the attribution of lies in scholarship; and those who resort to this practice risk summoning forth the raised eyebrow, if not the curled lip.² Nonetheless, we believe that in well-conducted exegesis attributions of desirable qualities, such as consistency and astuteness, can (and ought) sometimes be traded off against undesirable (but human) qualities such as lying. The latter can come to the rescue of the former. We further believe that the attribution of lies can pay dividends in historical scholarship no matter how strongly the

author being interpreted condemns lying. Accordingly, we argue that our approach works even when applied to one of the strongest condemners of lying: Immanuel Kant.

We discuss several instances in which we believe Kant lies, but our central case is the following. There are substantive changes between the first and second editions of the *Critique of Pure Reason*—changes which would be obvious if the author had not insisted in the second edition that such changes as he makes there are not substantive but merely concern the ‘manner of exposition’ (B xxxvii-B xxxviii). We contend that instead of interpreting Kant as oblivious to the substantive nature of the differences between the two editions, it is more charitable to conclude that he lied. Flagging Kant’s ‘no substantive changes’ assertion as a lie enables readers to trust their impression of substantive differences between the editions without compromising their confidence in Kant’s astuteness or logical acumen.³

This strategy has the merit of preserving the rationality of both author and reader. Nonetheless, it faces a prima facie objection deriving from Kant’s own ethics. If Kant really did regard all lies as grave misdeeds, then he would not have told anything that amounted (by his own lights) to a lie anywhere in his scholarly work – and certainly not in a manner as readily discernible as we claim! Perhaps what *we* see as lies Kant’s did not count as such. To pre-empt this objection, we examine Kant’s account of lying. We argue that the apparent lies he tells do indeed amount to lies even by his own lights. To defend our attribution of lying to Kant we discuss factors that would have given him a powerful incentive to lie. These factors, when combined with biography, dictate a choice: Either we attribute to Kant an ordinary level of mendacity or we saddle him with an extraordinary level of obtuseness (because we treat him as blind to the changes). For reasons of overall charity, we prefer the first alternative.

1. Shifting allegiances: Kant’s conception of lying in its historical context

In *The Groundwork of the Metaphysics of Morals* (1785) Kant took himself to have deduced a universal prohibition against lying from reason itself (4: 389). He defended his maxim ‘Never lie’ against putative counterexamples that chasten all but religious absolutists. Even Augustine and Aquinas couple their own ‘Never lie’ injunctions with a mitigating distinction between venial and mortal lies.⁴ In the view of these Saints, lying is always forbidden but violations are often minor offenses. In strident contrast, Kant’s only concession to practicality is taxonomic. Sometimes (but not always) he denies that polite lies, joking lies, and ‘necessary lies’ [*Notlüge*] are, properly speaking, lies.⁵ Just as there are misnomers in biological taxonomy (koala bears are not bears, glass snakes are neither glass nor snakes), there are misnomers in moral taxonomy. For example, the taxonomic never-liar courteously declares: ‘White lies are no more lies than white chocolates are chocolates!’. Since they override surface grammar such re-classifications need defending. The never-liar does so by specifying a necessary condition for lying that is not met in these cases – as it might be, malevolence.

What is the root of all sin? Lying, answers Augustine, because the speaker turns away from Truth which is identical to God (Puffer 2016, 686-688). ‘A lie in the mouth kills the soul’ (Wisdom 1:11). Initially, Augustine makes an allowance for spiritual immaturity (Puffer 2016, 697). A boy is praised for lying as rarely as he does just as he is praised for making so few grammatical errors. Some lies are quite forgivable. For instance, jokes are ‘lies to which no great blame is attached, and yet they are not completely without blame.’ (Augustine 2004, 5.7).

Three years later, in ‘On Lying’, Augustine withdraws his developmental allowance and ejects jokes from the spectrum of lies. A speaker lies only if he intends the addressee to believe the lie; and the joker ‘has in mind no deceit’ (Augustine 1952, 1.2).

Aquinas agrees that the normal trajectory of a lie is deception. But deception is no more necessary for lying than aiming at a target is necessary for shooting an arrow.⁶ Aquinas therefore restores jokes – or such of them as make assertions – to their original status as lies.

Following Aristotle's teleology, Aquinas claims that the natural function of the language organ is to communicate truth. The tongue-in-cheek speaker perverts this function by intentionally asserting something he believes to be false: '[A]s words are naturally signs of intellectual acts, it is unnatural and undue for anyone to signify by words something that is not in his mind' (1978, qu. 110, art. 3).

Since Aquinas counts as a lie any assertion of what one believes to be false, on his view Kant sometimes lied when he told jokes.⁷ Even Augustine will be tempted to classify some jokes (as *Kant* conceives of these jokes) as lies. For Kant assigns a structural role to deception in jokes. 'It is noteworthy', he says 'that in all such cases the joke must contain something that can deceive us for a moment. That is why, when the illusion vanishes, transformed into nothing, the mind looks at the illusion once more in order to give it another try and so by rapid succession of tension and relaxation is bounced back and made to sway.' (5: 332). By Aquinas's lights, Kant was encouraging lying when he advised guests at dinner parties to follow serious conversation with playful wit and jests (Kant 2006, 181 [281]). Spared from condemnation are jokes that turn on double-truths, such as Kant's joke about 'the 'lack of eminence' in his backside' (Kuehn 2001, p. 418).

Some jokes plainly are lies. Catholic moralists repeat an anecdote, probably apocryphal, of Thomas Aquinas who was dubbed by classmates 'the dumb ox'. The nickname seemed apt because he was a big and quiet. Told 'Look! There is a flying ox!', Aquinas lumbers to the window. Scanning the heavens, he is disappointed by the absence of a flying ox. Classmates laugh at the dumb ox's gullibility. Aquinas retorts that he is more surprised that a Catholic would lie than

an ox fly. The fact that Aquinas's retort would itself count as a lie by his own lights provides further reason to think the story apocryphal.

In his lectures on Anthropology, Kant tells a variant of this story featuring Peter Abelard (25: 1272). He concludes that even the 'cleverest minds' can be deceived. Axel Gelfert identifies two Kantian principles that may have figured as premises for Kant's conclusion: First, you are entitled to accept another's word on matters to which you lack empirical access (Gelfert 2006, pp. 633-636). Second, you must respect the experience of others as much as your own and so you have a duty believe *a posteriori* testimony (Kant 1900, XVI).

For those who require an intention to deceive, the charge of lying gains traction for the deadpan delivery favored by witty lecturers. The stone-faced Kant rarely laughed, leaving students unsure of whether a remark was serious or dry humor (Kuehn 2001, p. 106). Some students overcame Kant's suppression of facial cues. Others did not.

If Aquinas is right that lying is asserting what you believe to be false, Kant also lied when he misspoke.⁸ In the first *Critique*, for example, he (quite possibly) lies in this (not especially blameworthy) way by asserting about the term 'nichtsterblich' (non-mortal) something he really wants to say about the term 'nicht sterblich' (not mortal) (A 574/B 602) (The former term conveys predicate negation, the latter copula negation). When he makes this assertion—assuming this is not a type-setter's error—he says something about the former term which is, by his own lights, false: namely, that 'nichtsterblich' leaves all content untouched (ibid). Typically, when mis-speakers are confronted with such discrepancies, they experience mild embarrassment and apologize for having asserted what they do not believe.

Such a fussy attribution of lying to Kant might well seem childish. And in the most literal sense it is. For young children do indeed follow Aquinas in treating lying as asserting something one believes to be false. They even measure the gravity of lies by how far they depart from the

truth. They suppose that lies that depart greatly from the truth deserve more punishment than lies that stay close to the truth. Reporting that one saw a dog as big as a cow is naughtier than telling the more plausible lie that one received good marks when the teacher gave no marks at all (Piaget 1966, p. 152).

Around the age of ten, children convert to Saint Augustine. The misspeaker did not lie because he did not intend to deceive. Maturation involves reconceptualization of lies as a species of deception. Measured by these standards, Kant immatures with age. He starts out following Augustine in demanding the intention to deceive, arguing that one's false declaration is not a lie when one knows that one cannot deceive (27: 447). This remains his position for most of his career, though traces of Aquinas begin to appear in 1791 (8: 267, see below).⁹ The 90s remain a troubled decade, during which Kant accumulates commitments that clash with the Augustinian picture. Finally, he (tacitly) converts to Saint Aquinas in his 1797 essay 'On a Supposed Right to Lie from Philanthropy' (SRL).

The story of Kant's conversion is best told backwards. In SRL, the late-career Kant defines a lie 'as **an intentionally untrue declaration to another**' (8: 426). This definition makes clear that at this stage he conceives of lying as an intentional act. But, rather strikingly, there is no demand that the liar must intend *to deceive*.¹⁰ Nor is that requirement mentioned in Kant's *characterization* of a lie in the ethical sense in the *Metaphysics of Morals* (MM), a work which appeared one month before SRL in August 1797.¹¹ Nonetheless, the requirement *is* present at this stage, for it is mentioned in the related discussion at MM 6: 430. 'A lie', Kant tells us there, 'requires a second person whom one intends to deceive'. The discrepancy in MM between definition and commentary is, we believe, a symptom of a tension. In this work Kant wishes to retain the requirement of an intention to deceive but his views on the phenomenon of 'inner lies' create counter-pressure. He simply cannot see how one can deceive oneself:

It is easy to show that the human being is actually guilty of many inner lies, but it seems more difficult to explain how they are possible; for a lie requires a second person whom one intends to deceive, whereas to deceive oneself on purpose seems to contain a contradiction (MM 6: 430).

It compounds the tension that Kant is, at this stage, already attracted to Aquinas's picture. Some six years before encountering the problem about inner lies he flirts with Aquinas's account of lying (substituting 'truthfulness' for Aquinas's 'formal truth'):

One cannot always stand by the **truth** of what one says to oneself or to another (for one can be mistaken); however, one can and must stand by the **truthfulness** of one's declaration or confession, because one has immediate consciousness of this. For in the first instance we compare what we say with the object in a logical judgment (through the understanding), whereas in the second instance, where we declare what we hold as true, we compare what we say with the subject (before conscience). If we make our declaration with respect to the former without being conscious of the latter, then we lie, since we pretend something else than what we are conscious of. (*On the miscarriage of all philosophical trials in theodicy*; 8: 267; 1791)

Here Kant insists upon the Thomistic sufficient condition for lying without mentioning the countervailing Augustinian necessary condition. As we have seen, the tension in MM is short-lived: it is resolved in SRL within a month and it is resolved in favor of Aquinas. What might have

prompted Kant's final break with Augustine? One possibility, of course, is his worry about the coherence of self-deception. But another possibility is simply that he had reflected more on the phenomenon of inner lies made, specifically, to God. In MM he touches on this point without developing it.

Someone tells an inner lie, for example, if he professes belief in a future judge of the world, although he really finds no such belief within himself but persuades himself that it could do no harm and might even be useful to profess in his thoughts to one who scrutinizes hearts a belief in such a judge, in order to win his favor in case he should exist. (6: 430)

For Kant God is both the universal scrutinizer of hearts (*Herzenskundiger*) and at the same time the highest judge (7: 9-10; 21: 147). One lies to God if one inwardly professes one's faith in a future highest judge while not finding such a belief in one's heart. Since one cannot intend to deceive God, the Augustinian necessary condition must be dropped.

The weight Kant attached to these points must have increased between MM and SRL. Did he come to brood more deeply on the possibility that he had lied to God as he more nearly approached the day of his final judgment?¹² Did he become more sceptical about the possibility of self-deception? It is hard to say, but each of these sources of discomfiture would help to explain his final decisive break with Augustine. And that break *is* decisive; for in SRL there is, in the first place, no trace of the requirement that one should intend to deceive and, in the second, the possibility of lying to oneself is *definitionally* excluded: a lie is now an intentionally untrue declaration *to another* (8: 426).

The hypothesis that Kant should have changed his definition of 'lie' (in the ethical sense of the term) receives further support from the fact that it explains his differential attitude toward the

‘throat grabber’ and ‘murder at the door’ cases. In spite of the many parallels between the two cases, the first—at least according to the note-taker, Collins— is *not* a lie since the one making the rough inquiry ‘knows that I will withhold the information’ and so both I and he know he cannot be deceived’ (27: 447)¹³. And yet the second *is* a lie even though the murderer at the door has an equally strong reason to expect that I will withhold information in this case too. In ‘murderer at the door’, the lack of an intention to deceive does not, as it had before, serve to downgrade the lie to a mere untruth. One simple explanation of this difference is that Kant has dropped the Augustinian requirement of an intention to deceive. In dropping it, he also loses his rationale for declining to classify throat-grabber fabrication as a lie.

Augustine’s opinion about what is a genuine lie narrows as he tries to make ‘Never lie’ more practical to his widening audience. At the dawn of his intention-based theory, Augustine [1968, 233-4 (10.30)] is committed to counting the duping delight of a dreamt lie as a violation of ‘Thou shalt not bear false witness’ just as he counts wet dreams of fornication as violations of ‘Thou shalt not commit adultery’ (extrapolating from Jesus’s extrapolation of adultery to lustful-looking in Matthew 5: 27-28). The *wish* to lie is a lie. Analogy: one may lie in a letter that is never received by the intended victim of the deception. It is the thought that counts!

The intent to deceive requirement does classify some apparent lies as non-lies. Suppose you know your friend is convinced that you will answer mistakenly when asked whether *p*. Concerned about her safety, and wanting to cause her to believe the truth, you answer that *p* contrary to your belief that not *p*. In this case your intention to deceive her about whether you believe that *p* serves as a means of getting her to believe what you really believe. Since your assertion that *p* was not intended to deceive her into believing that *p*, you did not lie! This is good news for faculty who feigned belief in Kant’s excuses. ‘Even when his ceremonial position as

rector of the University of Königsberg required him to attend religious services, he always declined, reporting that he was “indisposed” (Bird 2006, p. 26).

The more Augustine explores lying, the more he judges the topic ‘Full of dark corners, with many cavernous windings’ (DM I, 1). He suspects he has made undetected mistakes when sorting lies from their look-alikes. This humility accords with his belief that the human intellect naturally develops over time. Augustine sets an example to all scholars by publishing his voluminous ‘Retractationes’.

2. Theories of Error as Sources of Prevarication

In his *Meditations*, René Descartes claimed to have eliminated all possibility of error through his method of doubt. In line with this meta-assessment, he continued to affirm each one of his conclusions in the teeth of his waning confidence in substance dualism (Machamer and McGuire 2009). As his confidence waned, Descartes inched closer to asserting things he did not believe, hence to lying –at least on Aquinas’s conception of lying. Descartes’s confidence in his methodological insulation from error thus served as a spur to mendacity.

The risk of an inconsistent combination of views grows exponentially with each assertion. Yet a scholar content to apply this scale effect to others will sometimes seek protection for his own book. In a letter to Johann August Becker (March 3, 1854) Arthur Schopenhauer wrote: ‘To seek contradictions in me is completely idle: all is from one gush.’.

Schopenhauer’s image of wholesale or unitary assertion may have been influenced by Kant’s depiction of his critical philosophy as prompted by the recollection of David Hume as having roused him from his dogmatic slumber. Upon awakening the dreamer’s delusions melt

away. This wholesale withdrawal of assent from falsehood might possibly have struck Schopenhauer as a model, when suitably inverted, for his gush of insight.

In fact, however, Kant's confidence in the *Critique of Pure Reason* is more akin to the 'Aha!' that comes from dialing the correct combination of a lock. Every element must be in perfect alignment!

Pure speculative reason, and Kant's system insofar as it enshrines it, is an organic whole (B xxxvii-xxxviii). Like an organism, the *Critique* is a self-sustaining system in which the parts exist for the whole and the whole for the parts (ibid.).¹⁴ Kant believes that owing to the systematic arrangement of the first *Critique's* parts and its plan-guided construction, even the slightest flaw—whether an error or an omission—would have betrayed itself at the time of its composition through contradictions, not just in the system but in pure reason itself (B xxxviii). Kant's failure to detect such contradictions gives him reason to hope that his system 'will continue to sustain itself in its unalterable state' (ibid.). At any rate, it is clear that he regards it as irrefutable (B xliii) and 'not in the slightest danger' (B xliv).

We are told that insofar as the second edition does make changes they affect only 'the manner of exposition' [*Darstellungsart*] (B xxxviii-B xxxix). They are needed only to rectify an obscuring 'prolixity' in certain parts of the work (4: 381). They are prompted not by the identification of mistakes but only by 'misunderstanding[s] by competent and impartial critics' (B xli). Crucially, Kant tells us that in 'the propositions themselves and their proofs, and also in the form and completeness of the plan, I have found nothing to alter' (B xxxvii).

Nonetheless, while tidying up for the second edition, Kant must have noticed that he was correcting substantive errors. Since this did not fit with his vision of his system as a flawless, self-sustaining organism (B xxxvii-B xxxviii), he falsely presents the changes as cosmetic. Here is an inventory of these substantive changes:¹⁵

First, there is an important change in the Analogies of Experience. Kant is forced to make a substantive change to his formulation of principle of persistence of substance, which, in the first edition, inadvertently carries the dogmatic consequence that the soul is permanent. What Kant came to realize in the period between the two editions is that in its *A*-edition formulation this principle, which he regards as stating a truth known a priori, applies as much to the object of inner sense—the thinking I as it appears to itself—as to the objects of outer sense. As Ralf Bader has convincingly argued, he thus came to believe that in the first edition he was, contrary to his intentions, dogmatically committed to the existence of a permanent object of inner sense—a dogmatic commitment that betrays itself (albeit in watered-down form) in the *Prolegomena* (Proops 2021, 118–122). It seems to have been this realization that prompted Kant to reformulate the principle of the first analogy in the *B* edition so that it applied only to the objects of *outer* sense – i.e., to bodies (Bader 2017, 213–15, Proops *ibid.*).

The exact nature of the reformulation is not readily apparent from the wording of the two formulations, but, as Bader observes, it becomes clear from Kant’s accompanying remarks in various reflections. This point is worth developing. First let’s review the two formulations in question:

The principle of the first analogy in the A-edition:

All appearances contain the permanent [*das Beharrliche*] (**substance**) as the object itself, and the changeable [*das Wandelbare*] as its mere determination, i.e., a way in which the object exists. (A 182)

The principle of the first analogy in the B edition:

In all change of appearances substance persists, and its quantum in nature is neither increased nor decreased. (B 224)

The first formulation, since it says that *all* appearances (hence both outer and inner) contain the permanent, clearly commits Kant to the permanence of the object of inner sense (i.e., the thinking I - it is this that appears (to itself) in inner sense). The second, formulation drops the containment claim and replaces it with a claim that entails the permanence of such substance as one finds in *nature*. This opens the possibility that in this second formulation the phrase 'all change of appearances' is supposed to apply only to such changes as occur in *nature*. As Ralf Bader has observed, certain reflections collectively suggest that this is indeed so, and, further, that 'nature' in the present context means the realm of change specifically in *space* - i.e., the realm of bodily motions as described by physics, something that Kant sometimes refers to as 'bodily nature' (28: 875)).

Thus, speaking of the proof of the principle of the first Analogy at A 182, Kant says 'Here the proof must be so conducted that it applies only to substances as **phenomena** of outer sense' (Reflection 80, 23: 30) (The phenomena of outer sense are found in space). In what seems to be a comment on this same principle, Kant says: 'Here it must be shown that this proposition does not pertain to any other substances than those whose alteration is effected only through moving causes, [which alteration thus] consists in motion, hence in the alteration of relations' (Reflection 77, 23: 30). But those items whose alteration is effected only through *moving* causes are, for Kant, bodies. In a third reflection Kant concludes that for these reasons 'my [own] permanence is not proven' (Reflection 83, 23: 31)

If, as seems plausible, these remarks are comments on changes that Kant felt he needed to make between the editions, they collectively suggest that in making those changes Kant divested himself of an unintended and unwanted dogmatic commitment to the permanence of the self. Plainly, however, the excision from the *Critique of Pure Reason* of an unwanted dogmatic commitment must count as a substantive change.

The second major change amounts to a change in the terms of Kant's criticism of the first paralogism. In the first edition the argument he presents as the first paralogism is treated as a sound argument which establishes too little to be of interest, but which is nonetheless passed off as establishing something interesting. It is treated, in other words, as an *ignoratio elenchi* (see A 350-1, A 400, and compare *Prolegomena*, first paragraph of section 47). In the second, however, Kant diagnoses the first paralogism as a *sophisma figurae dictionis*: a fallacious syllogism infected with equivocation on its middle term (see B 411-12, and the footnote on this text). The terms of his diagnosis of this central error in dogmatic rational psychology thus alter dramatically between the two editions.

The third major change is a change in the terms of Kant's criticism of the second paralogism. In this even more consequential case, he shifts his diagnosis not away from the charge of an *ignoratio elenchi* but towards it. Whereas in the first edition he criticizes the argument of the second paralogism on the ground that a crucial premise of the argument is not known to be true (A 353), in the second, abandoning this claim, he charges that the argument establishes too little to be of interest to the dogmatic metaphysician (B 407-8). The crucial premise in question is the claim that a thought cannot be thought by a plurality of thinkers in which each member thinks a distinct component of the thought. In the first edition, this claim is treated as not known to be true (A 352-3), but in the second it is taken to be known to be true. Kant adjudges the claim true when he says: "That the **I** of apperception, and hence in all thought, is a **singular** that cannot be resolved into a

plurality of subjects and therefore designates a logically simple subject—this lies already in the concept of thought and hence is an analytic proposition.” (B 407-8).

He means that because the components of a thought are thought of as unified in the thought, one cannot suppose that a plurality of thinkers severally thinking the components of a thought jointly pull off the thinking of the whole, and, accordingly, one cannot understand the thinker of any thought to be a plurality or group mind.¹⁶ (A *logically* ‘simple’ subject is something that is one thing rather than many things: it is a *single* subject rather than a subject lacking proper parts). Kant, moreover, continues to adhere to the view endorsed at B 407-8 in the decade following the publication of the first *Critique*. In the *Real Progress* of 1793, for example, he says:

The unity of consciousness, which must necessarily be met with in every cognition ... makes it impossible that representations distributed among many subjects should constitute unity of thought.¹⁷

At least from the *B* edition onwards, then, Kant is prepared to affirm something that in the *A* edition he’d claimed we could never have grounds for affirming.

In the *B* edition, accordingly, he changes his criticism. He now treats the second paralogism as designed to show that the soul is simple not in the sense of being a non-plural subject of thought (which might nonetheless inhere in another thing), but rather in the sense of being *one substance* (a necessarily non-inhering subject of inherence) (B 288, B 407, 4: 503, 8: 225). And he changes his objection accordingly. His complaint is now that the argument’s conclusion, namely, that the soul is a non-plural *subject*, though *successfully* established, gets inflated to the claim that the soul is a single *substance*.¹⁸ The argument that Kant regards as sound to the uninflated conclusion runs as follows:

Anything whose action can never be regarded as the concurrence of many acting things is **simple** (that is, not a group actor).

The soul, or thinking I, is something whose action can never be regarded as the concurrence of many acting things.

The soul, or thinking I, is simple (that is, not a group actor – hence, given that the action of the thinking I is to think, not a group *mind*).

Kant's second-edition criticism of this argument, which implicitly presents the argument as an *ignoratio elenchi*, occurs in the continuation of the passage already cited from B 407-8. Kant says:

That the I of apperception, consequently in every thought, is a **single thing** [*ein Singular*] that cannot be resolved [*aufgelöset*] into a plurality of subjects, and consequently that it denotes a logically simple subject, lies already in the concept of thinking and is therefore an analytic proposition; but this does not mean that the thinking I is a simple **substance**, which would be a synthetic proposition. (B 407-8).

In this context, to 'resolve' the 'I' of apperception into a plurality of subjects means to treat each of my thoughts as the non-emergent product of the joint activity of a plurality of subjects of representation. The first part of this remark, accordingly, amounts to the claim that the minor premise of the argument just stated is an analytic truth. Since Kant adds that this claim entails that the 'I' of apperception *denotes* a logically simple subject, it is clear that 'logical simplicity' is now a stronger notion than it had been in the *A* edition: whereas in the *A* edition logical simplicity had been a property of representations, it is now treated as a property of what is denoted by a

representation. The second part of the remark warns us that we cannot infer from this conclusion the stronger conclusion that the thinking I is a simple substance, and here ‘simple substance’ means a *single* non-inhering subject of inherence (see our note 18 above).

In the second edition, then, Kant’s criticism becomes the charge that the second paralogism is an *ignoratio elenchi*: the paralogist commits the sophism of misconstruing (and exaggerating) what has been established. This charge of inflation of a proven claim is a stark change from the first edition, where the criticism had taken the form of the charge that the argument did not establish its purported conclusion simply because its minor premise was not known to be true (A 352-3).

This is a change in the terms of Kant’s criticism, but it also involves a change in (and strengthening of) the claim that Kant presents as the rational theologian’s target thesis. In the first edition the rational psychologist is depicted as running the argument of the second paralogism in the attempt to prove that the thinking I is a simple (i.e., single) *subject of thought*. In the second, realizing that this goal is in fact *attainable* (because the thesis in question is now regarded as an analytic truth) Kant raises the stakes. Now the rational psychologist is depicted as attempting to prove that the thinking I is a single *substance* (in Kant’s technical sense of that term, namely, a necessarily non-inhering subject of inherence). Since the second claim is stronger than the first, Kant has moved the goalposts. Changing the metaphor, he has set the bar higher for success in the second edition than he had in the first. Worse, the change is unacknowledged: he has failed to report this dramatic alteration in the Dialectic’s plan.

The fourth major change between the two editions concerns the fourth paralogism. This is again a change in the content of the supposedly unwarranted thesis that Kant attributes to the dogmatic rational psychologist. In the second edition what had been an *epistemic* thesis, namely: ‘The existence of objects of the outer senses is doubtful’ (i.e., not known with certainty) (A 367) is

replaced by a *metaphysical* thesis, namely: 'I can exist merely as a thinking being' (i.e., without a body) (B 409). Accompanying this striking change of thesis is a change in the nature of the mistake that the paralogist is said to be making. Instead of making an epistemic claim that Kant regards as *false*, the paralogist is taken to be making a metaphysical claim that Kant regards as *not known (on theoretical grounds) to be true*.

There are other substantive changes between the two editions—many of them occurring in the murky swamp of the transcendental deduction—but these relatively pellucid examples already make clear that certain theses and proofs were indeed substantively revised. The *plan* of the work also changed insofar as Kant came in the second edition to depict what had been the target conclusion of the second paralogism as provable after all, while also adding the Refutation of Idealism—though this, as we shall see, is a change that Kant actually admits. The details of some of the changes we discuss will no doubt be controversial, but in the light of the foregoing discussion it should be plain that the weak claim that *some* substantive changes were made, which is all we need for our purposes, should not be controversial.

Kant seems to have realized that he wasn't going to get away with claiming that he'd made not the smallest addition or change in the 'ground of proof' of any proposition, for he admits, in a somewhat sheepish footnote, that he'd in fact done both things when he'd added the 'Refutation of Idealism' to the second edition.

Only one of my alterations could I call, properly speaking, an addition, and even it concerns only the kind of proof I offer. It consists in a new refutation of psychological *idealism*, and a strict proof (also, I believe, the only possible proof) of the objective reality of outer intuition. (note to B xxxix)

Since the new proof replaces the old one, Kant's 'only possible proof' claim tacitly acknowledges that the old proof of the objective reality of outer intuition had been incorrect. Since the new proof lies outside the paralogisms while the old proof had been given in the criticism of the fourth paralogism, the plan of the *Critique* changes along with this change in a 'ground of proof'.

Here we have a moment of partial candor (only partial because while the change of ground of proof is acknowledged, the change of plan is not). But Kant's admission is inadequate: our examples show that he also changed his account of the 'grounds of proof' of his claims that the first and second paralogisms are sophistical arguments, while also reformulating the target theses of the first, second, and fourth paralogisms. (Arguably, the target thesis of the third paralogism also changes because Kant changes his definition of a person, but this is not a point we will press here.) He also reinterpreted the principle of persistence of substance (envisaging commensurate changes to its associated proof).¹⁹

Nor can one downplay these changes as too modest to count as 'substantive' revisions from a suitably lofty perspective. For, by any measure of 'scholarly heft', they are equal in weight to the change admitted in the footnote. Indeed, a reinterpretation of a principle that had landed Kant in unwitting (and unwanted) dogmatism—the principle of persistence of substance—seems to be a bigger change than a reformulation of an argument for the objective reality of objects of outer intuition. Without the reinterpretation of the principle, after all, Kant would saddle himself with a theoretico-dogmatic commitment to the natural immortality of the soul. He would thus – rather disastrously – forfeit his claim to be a fully critical philosopher.

Might the changes in question be downplayed on other grounds? Might one claim that when Kant insists that he has not altered even 'the smallest part' of the *Critique* (B xxxviii) he actually intends to be speaking merely of certain claims that we might see as belonging to the book's essential core, a core comprising only such weighty highlights as the transcendental

deduction, the proofs of transcendental idealism, the positing of a priori synthetic cognition, the idea that we have a priori intuitions of space and time? Might we, that is to say, read Kant as exempting from his 'no changes' claim the changes we have just described?

The answer is 'no'. The changes in the paralogisms simply cannot be treated as disqualified by their smallness from what Kant describes as the *Critique's* 'smallest parts'. For Kant specifically mentions the paralogisms as a part of the work that contains changes only in the manner of exposition (B xxxviii). In the *Prolegomena* he mentions them again as incorporating changes designed to remove only a certain obscuring 'prolixity' (4: 381). These claims are flatly false, and they are known to be false by the author of the *B*-edition. They are, accordingly, lies. Nor could Kant's abandonment of his commitment to the natural immortality of the soul count as a 'small' change. Natural immortality, after all, entails for Kant one of the two 'cardinal' propositions of theoretically-grounded speculative dogmatic metaphysics, namely, a post-mortem existence for human beings (A 741-2/B 769-70; compare 18: 273). If the Kant of the first edition had supposed we could know the soul to be naturally immortal on theoretical grounds, he would have been a theoretico-dogmatic speculative metaphysician. Of course, this is the last thing he aspired to be.

There is no getting away from the fact that from Kant's point of view the first edition contains significant and substantive mistakes. And yet Kant's organicism prevented him from acknowledging these mistakes. According to that conception, everything in the structure of the *Critique* is an 'organ'. Everything in the system 'is there for the sake of each member, and each individual member is there for the sake of all' (B xxxvii-B xxxviii). In consequence, 'even the slightest defect' in the system whether it be 'a mistake (error) or an omission' ramifies through the whole, rendering it inconsistent (ibid). Less dramatically, a work that contained substantive errors would be incomplete in its representation of the truth. But for Kant a critique of pure

reason 'is never trustworthy unless it is **entirely complete** down to the least elements of pure reason' (4: 263). In this domain one must settle 'either **all** or **nothing**' (ibid.). It follows that if any part of the first edition had been mistaken, the whole, being an incomplete representation of the truth, would have been, at the very least, 'untrustworthy'. But then the parts of the first edition that were retained in the second would have themselves been untrustworthy. Accordingly, to admit his mistakes in the preface to the second edition would have been to admit that, for all he knew to the contrary, the work he was now presenting to the world might be riddled with error.

By constructing a theory to certify the correctness of his book Kant had made a rod for his own back. With the stakes thus raised, he acquired a powerful motivation to lie about the changes he had made. The motivation, moreover, soon took on a human aspect. Kant had been stung by Feder and Garve's mean-spirited (not to mention uncomprehending) review of the first edition of the first *Critique*, and especially by the charge that his system amounted to a 'higher' idealism. Aggrieved, he inserted into the *Prolegomena* a somewhat tetchy response alleging grave misreadings. His dander up, Kant issued a swashbuckling challenge to his opponents to show that even one of his arguments in the antinomies was invalid (4: 379). Given this combative response, he could hardly have made the following admission in the second edition:

In all candor, the first edition contained substantive errors and was therefore, by its own organicist lights, inconsistent and so untrustworthy throughout. Here's a new version remedying the flaws that I have discovered. Readers should alert me to any further flaws.

In addition to making the *Prolegomena's* challenge look foolish, such an acknowledgement of fallibility would have rendered the question whether he'd been misunderstood somewhat moot. Worse, it would have exposed the first edition's exercise in self-certification to public derision.

Kant lied because he had painted himself into a corner and knew that his less sympathetic critics must have been poised to seize gleefully on any concession, especially one that entailed that the first edition had been inconsistent.

3. When your theory helps you lie

‘Scholarly lie’ is not a misnomer for Kant. To characterize a substantive correction as a mere clarification is, by Kant’s own taxonomy, to lie.

Of wider significance for scholars are lies told because their theories imply those utterances are not lies. The scholar’s scruples against lying are applied only to contemplated actions that the scholar classifies as a lie. Most scholars do not think they lie when they use a translation without citing the translator. But some of the uncited translators do regard this as plagiarism; the author is passing off the translation as their own.

We disagree. Uncited translation is too common a practice to raise an expectation that the translator be cited. The indignant translator replies that widespread lying is still lying. And they have a point. Nowadays, virtually all downloaders of software lie when required to declare they have read the service contract. These routine lies are not meaningless formalities – as is evident from the great care with which contracts get modified by the vendors’ lawyers. The fact that downloaders sincerely deny they are lying does not override the attribution of lying by vendors.

Nonetheless, although we are persuaded by the vendors, we are not, at the end of the day, persuaded by the indignant translators. All the same, we do enthusiastically endorse the translators’ general point that some scholars lie because they fail to count some lies as lies. Archimedes, annoyed with those who filched results, sent colleagues theorems without proofs and included two false theorems. ‘... so that those who claim to discover everything, but produce no proofs of the same, may be confuted as having pretended to discover the impossible.’ (Preface to ‘On Spirals’

from Heath. 1912, p. 151). He notes the obvious inconsistency in one of the falsehoods. Perhaps Archimedes thought lies are permissible when they are designed to expose a few fakers (albeit at the cost of deceiving the majority of readers). A more pertinent hypothesis, however, is that he, and perhaps also Kant, did not regard these falsehoods as lies (Sorensen 2022). After all, Archimedes's audience is in a position to deduce the falsehood. Since theorems cannot be learned by testimony, mathematical readers are obliged to be vigilant, repeating only what they understand through demonstration. The cause of their belief in pseudo-theorems is their dereliction of duty, not the craftiness of the mathematician who exposes their negligence, irresponsibility, and – in some cases – piracy.

In the case of errors concerning a priori matters, such as the erroneous belief that a 1/4 pound hamburger weighs more than 1/3 pound hamburger, blame the deceived not the deceiver! This imperative fits Kant's (and Descartes's) explanation of why *a posteriori* lying is worse than telling misleading truths – at least according to Alasdair MacIntyre (1995, p. 337) and Jonathan Adler (1997, p. 144). Although the intention to deceive is the same, those deceived by a misleading truth must make a hasty inference beyond what was said. This recklessness weakens the hearer's ability to discern whether he is acting from a good will. He is also apt to infect others with his negligently formed beliefs.

Many speakers are reluctant to count as lies the falsehoods used to detect and prevent lies. Such a crafty use of falsehoods is institutionalized in the technique of planting tell-tale falsehoods in scholarly publications in order to expose the copy-cat. Compilers of mathematical tables insert a few slight errors `as a trap for would-be plagiarists` (L. J. Comrie, quoted by Bowden 1953, p. 4).

To deter copyright infringement, cartographers insert fictitious streets – so-called 'trap streets' – into maps. The cartographers may not regard these fabricated lines – or such of them as remain anonymous – as lies because the representation is pictorial rather than discursive. Kant

requires all lies to be sentence-like. For him ‘unstated lie’ is a contradiction in terms; for a lie is an intentional *declaration* to another (8: 426). Kant’s exclusion of lying by omission is a relief to scholars who engage in ‘the historian’s revenge’. They refrain from acknowledging enemies in a chronicle that might make memorable the enemies’ contributions.

Not excused on Kantian principles are medical historians who announce they will use false names and confound circumstances to protect the privacy of patients (Mukherjee 2010, p. xiv). The reader may take himself to have given the medical historian permission to lie for a good purpose. But Kant still forbids the particular lies.

Scholars may honestly mistake a lie for a non-lie. After all, there is much reasonable disagreement over the necessary conditions for a lie (let alone the full definition). Must lies be *false* (Carson 2010, p. 17)? Must lies be *known* to be false?²⁰ Must a lie be *understood* by the addressee? A hundred such questions could constitute a challenging examination for any expert. The scholar’s theory of lying will help her answer the hundred questions more consistently and completely. Thanks to her theory, some of the answers will correct myths about what is a lie. However, the theory will also lead to inaccurate innovations – especially if the theorist condemns all lying.

Some questions about lies cannot be reliably answered because ‘lie’ is vague. When asked your height, you may honestly answer ‘My height is n centimeters’. But you could have also honestly answered $n + 1$ centimeters. Continued addition of centimeters leads to answers that would be lies. But there is no knowable point at which the honest answers end and the lies begin.

Unclearly about authorship creates many borderline cases of lying. Post-dating was common among publishers who increased sales by issuing a product that appears fresher than it was. (Kant’s *The Only Possible Argument in Support of a Demonstration of the Existence of God* bears a 1763 publication date, but was in fact published mid-December of 1762.) The practice is

still common. More puzzling is the practice of including on the copyright page: ‘The moral rights of the author have been asserted’. Is this the author’s first assertion? Few authors notice the sentence and even fewer know what it means. Contemporary never-liars do not try to err on the side of safety by remonstrating with the publisher.

These borderline cases of lying contrast with a clear case in Kant’s compilation of previous work, *The Conflict of the Faculties*. Kant claims he was 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). Yet only eight years earlier, Kant had indeed failed to ‘sign his name’ to an anonymous work, namely: ‘On the propensity to fanaticism and the means to oppose it’ (1790). Moreover, he had taken the same precaution on *nine* previous occasions. They are listed on Steven Naragon’s ‘Kant in the Classroom’ website:

<https://users.manchester.edu/facstaff/ssnaragon/kant/Helps/KantsWritings.htm>).

Some popularly attributed lies border on conceptual incoherence. Kant says ‘One’s inner statements to oneself are statements, and, if other conditions are also met, can be “internal lies”’ (6: 429). Roderick Chisholm and Thomas Feehan cannot imagine how lying to oneself conforms to the requirement that the liar intends to deceive the addressee.²¹ Nevertheless, they are persuaded by another controversial claim occurring in the ‘throat grabber’ passage we alluded to earlier:

Our third case is proposed by Kant. ‘I may make a false statement (falsiloquium) when my purpose is to hide from another what is in my mind and when the latter can assume that such is my purpose, his own purpose being to make a wrong use of truth. Thus, for instance, if my enemy takes me by the throat and asks where I keep my money, I need not

tell him the truth, because he will abuse it; and my untruth is not a lie (mendacium) because the thief knows full well that I will not, if I can help it, tell him the truth and that he has no right to expect it of me.⁹²²

Chisholm and Feehan endorse Kant's pre-SRL opinion that the liar must invite the hearer to take the remark on faith. The requirement of invited trust might also explain Kant's toleration of bluffing (as it does for Carson 2010, p. 191). The practice of playing card games 'cultivates us, makes us even-tempered, and it teaches us to keep our emotions in check. In this way it can have an influence on our morality.' (*Anthropology*, 7: 275)

Conclusion

Members of the House of Commons are prohibited from lying but also prohibited from singling out a colleague as a liar. Accordingly, Stanley Baldwin once claimed to have caught Lloyd George telling the truth. Against Kant, such a quip would have fallen flat. Professors are not under the same pressure to be two-faced to appease rival coalitions.

When Kant was writing the *Critique of Pure Reason*, the only strains on his honesty emanated from first, a self-locking theory of error avoidance and, second, an ambition that his book be quickly recognized as a definitive breakthrough. Regrettably, but understandably, these strains proved too powerful for Kant's ordinarily fragile honesty to bear.

This is not to deny that other pressures exogenous to his masterwork later made their presence felt. Kant had to climb from the bottom of the social ladder. As he rose to become a regular participant at aristocratic dinner parties, Kant had to be diplomatic. As he turned from theoretical to practical philosophy, he faced far more pressure to prevaricate. In his essay

‘Perpetual Peace’, Kant characterizes people as natural born hypocrites. They keep each other in line by applying principles of justice self-interestedly. Instead of being genuinely committed, they make exceptions for themselves. The more power they acquire, the more corrupt they become. As Kant became influential, the more concern he came to have about his own commitment.

In the *Critique of Practical Reason* Kant approvingly reports intolerance of hypocrisy among the ancients. No one was entitled to call himself a philosopher, a lover of wisdom, ‘unless he could establish [philosophy’s] infallible effect on his own person as example’ (1956, 109 fn 5). Kant’s endorsement of this high standard makes biography immediately relevant. The first biography of Kant was written by Kant’s colleague and professor of medicine, Johann Daniel Metzger. Metzger knew Kant gave ill-informed medical advice. Metzger was also aware that earnest correspondents, such as Maria von Herbert, sought and received moral sedatives that applied maxims such as ‘Never lie’ (Langton 1992). Their credulity had made Kant credulous of his own status as the sage of Königsberg. According to Metzger. Kant said ‘he was leaving this world with a clean conscience, never having intentionally committed an injustice’ (Kuehn 2001, pp. 4-5). Metzger disbelieved Kant. True, Kant was a great philosopher. But he was only an average man, ‘neither good nor evil’. Metzger warned students not to emulate Kant. This provoked pastors to write hagiographic biographies (Kuehn 2001, 6-16). These are in the same genre as Mason Locke Weem’s *The Life of George Washington*. In reverend Weem’s fifth edition of 1806 the six-year old George answers his father’s query about a damaged cherry tree: ‘I cannot tell a lie...I did cut it with my hatchet.’. Kant did not lie nearly as much as the realtor, general, and politician, Washington. But neither did he lie much less than other professors of his era.

Unlike David Hume, Kant allows that judgments of duty can be (and indeed must be) motivating: ‘What I recognize immediately as a law for me’, he says, ‘I recognize with respect.’ (1959, 20-21). This leaves little room for weakness of the will. If someone really believes that lying

is always a grave misdeed, he will not lie as frequently as the average Prussian Professor. But Kant did.

Voltaire claimed to be universally indifferent to all metaphysical propositions. And yet he is a deist, a determinist, and defender of absolute space. Kant charitably renders partisan self-proclaimed indifferentists, such as Voltaire, consistent by denying the sincerity of their claims to universal indifference: they merely ‘try to feign’ indifference (A x).²³ Kant famously claimed to be universally opposed to lies. Taking a leaf out of his own book, we charitably render him consistent by denying the sincerity of his claim to be universally so opposed. In truth, he believed that lying is sometimes permissible. Kant, we believe, was acting on this permission when he denied making substantive changes to the first *Critique*.

Persistent over-generalizers pose a dilemma for the principle that people should be interpreted as moral by their own lights. Either we take them at their word, and attribute a big moral error, or we postulate a lie and attribute a smaller moral error. Since some error must be attributed, we charitably attribute the smaller error. Kant emerges as an opponent of lying who overstated his opposition. Many parents do the same to discourage children from lying. These parents have an empathic foothold on understanding Kant.²⁴

Ian Proops

The University of Texas at Austin

Roy Sorensen

The University of Texas at Austin

Arché Research Centre at the University of St Andrews in Scotland

Abbreviations:

MM: *The Metaphysics of Morals*, in Kant 1900-- (*Die Metaphysik der Sitten in zwei Teilen* (originally published, 1797); 2nd edition: *Rechtslehre*, 1798; *Tugendlehre*, 1803 in Kant 1900-- 6: 205-355, 373-493, translated by Mary J. Gregor in A. W. Wood and M. J. Gregor 1996.

SRL: 'On a Supposed Right to Lie from Philanthropy' ('Über ein vermeintes Recht, aus Menschenliebe zu lügen.' (translation in A. W. Wood and M. J. Gregor 1996) in Kant 1900--, 8: 425-30, originally published 1797.

References

Adler, Jonathan (1997). 'Lying, deceiving, or falsely implicating', *Journal of Philosophy*, 94, pp. 435-452.

Aquinas, Thomas, (1972). 'Question 110: Lying', in *Summa Theologiae* (Volume 41: *Virtues of Justice in the Human Community*), New York: McGraw-Hill.

Augustine, (1952) . 'On Lying,' M. S. Muldowney (trans.) (51-110), and 'Against Lying,' H. B. Jaffee (trans.) (pp. 121-179), in R. J. Deferrari (ed.) *Fathers of the Church* (Volume 16: *Treatises on Various Subjects*), New York: Fathers of the Church.

Augustine (1961). Confessions, R. S. Pine-Coffin (trans.) New York: Penguin Books.

Augustine (2004). 'Expositions of the Psalms'. Edited by John E. Rotelle, O.S.A. Translated by Boulding, Maria, O.S.B.. In The Works of Saint Augustine: A Translation for the 21st century. Vol. III/15-20, pp. 392-418. Hyde Park. NY.: New City Press.

Bader, Ralf (2017). 'The Refutation of Idealism', in James O'Shea (ed.), *Kant's Critique of Pure Reason: A Critical Guide* (Cambridge: Cambridge University Press, 2017), pp. 205-22.

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

Bowden, Bertram Vivian (1953). 'A Brief History of Computation.' In *Faster Than Thought*, edited by Bertram Vivian Bowden. London: Pitman Publishing.

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

Chisholm, Roderick., and Thomas Feehan, (1977). 'The intent to deceive', *Journal of Philosophy*, 74. pp. 143-159.

Clewis, Robert R. (2020). *Kant's humorous writings: an illustrated guide*. New York: Bloomsbury Academic.

Gelfert, Axel, (2006). 'Kant on Testimony', *British Journal for the History of Philosophy*, 14, pp. 627-652.

Heath, Thomas (1912). *The Works of Archimedes with the Methods of Archimedes* New York: Dover.

Johnson, Samuel: (1755, 1967.) *Dictionary of the Language*, 2 vols., New York: AMS Press.

Kant, Immanuel (1900-) *Gesammelte Schriften* Berlin: Koniglich-Preussichen Akademie der Wissenschaften zu Berlin (now De Gruyter).

_____ (1956). *Critique of Practical Reason and Other Writings in Moral Philosophy*, trans. Lewis White Beck, Chicago University Press.

_____ (1959). *Fundamental Principles of the Metaphysic of Ethics* tr. T. K. Abbot 10th ed. London.

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

_____ (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.

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

- _____ (2006). *Anthropology From a Pragmatic Point of View*. Cambridge University Press.
- Kuehn, Manfred (2001). *Kant: A Biography* Cambridge University Press.
- Langton, Rae (1992). 'Duty and Desolation'. *Philosophy* 67, pp. 481-505
- Machamer, Peter and J.E. McGuire (2009). *Descartes's Changing Mind*, Princeton University Press.
- MacIntyre, Alasdair (1995). 'Truthfulness, Lies, and Moral Philosophers: What Can We Learn from Mill and Kant?' in *The Tanner Lectures on Human Values*, 16, pp. 307-361. Salt Lake City: University of Utah Press.
- Mukherjee, Siddhartha. (2010). *The Emperor of All Maladies: A Biography of Cancer*, New York: Scribner.
- Piaget, Jean (1966). *The Moral Judgment of the Child* New York: The Free Press.
- Proops, Ian (2021). *The Fiery Test of Critique: a reading of Kant's Dialectic*, Oxford University Press.
- Puffer, Matthew (2016). 'Retracing Augustine's Ethics: Lying, Necessity, and the Image of God,' *Journal of Religious Ethics* 44:4, pp. 685-720.
- Sorensen, Roy (2022). 'Kant tell an *a priori* lie' in Laurence R. Horn (ed.) *From Lying to Perjury* De Gruyter, pp. 65-86.
- Wood, Allen (2011). 'Kant and the right to lie reviewed essay: On a supposed right to lie from philanthropy, by Immanuel Kant'. *Eidos: Revista de Filosofía de la Universidad Del Norte* 15, pp. 96-117.

¹ Quotations from Kant's writings are drawn from the Prussian Academy edition (Kant 1900--), the first *Critique* being cited by the A/B pagination. English translations are our own, though often indebted to standard translations listed in the Bibliography.

² A student of media studies might conjecture that this essay has been emboldened by the *New York Times's* attribution of a lie to Donald Trump in 2016 (namely, the assertion that Barack Obama was not born in the United States). The *Times's* executive editor, Dean Baquet, justified the attribution in a National Public Radio interview discussed by the *Washington Post* media analyst Callum Borchers: 'Why the New York Times decided it is now okay to call Donald Trump a liar' *Washington Post* (September 22, 2016). Other leading news outlets followed the precedent: 'Lies? The news media is starting to describe Trump's 'falsehoods' that way', *Washington Post*, Paul Farhi (June 5, 2019).

³ Kant does not use the exact phrasing: 'I have made no substantive changes', but, as we will see, he does make assertions whose content, taken collectively, are accurately summarized by such a claim. For example, he says that in 'the propositions themselves and their proofs, and also in the form and completeness of the plan, I have found nothing to alter' (B xxxvii).

⁴ Kant's discussion of mental reservation evinces familiarity with Augustine and Aquinas, at least through commentators such as Alexander Baumgarten. Kant writings contain no overt comparison of these two Saints on lying, such as that itemized by Christopher Tollefsen (2012). But historians are often struck by epistemological and metaphysical resemblances between Kant and the modern scholastics influenced by Aquinas.

⁵ See 27: 64 and compare R 6896, 1776 (19: 199) for the claim that 'necessary' or 'white' lies (*Notlüge*) are often not lies proper. Kant says that in the cases in question there is no breach of obligation. If, as seems likely, this means no obligation to a particular individual, we would have to take him to be speaking here specifically of lies in the juridical sense. Something that is not a lie in the juridical sense may still be a lie in the ethical sense, however. In the Collins notes of Kant's ethics lectures, he contrasts joking lies with real lies which are intended to be taken as true (Carson 2010, pp. 78-79; 27: 62). For the claim, made in these same notes, that an extorted known falsehood, asserted without an invitation to trust is not a lie (*mendacium*), see 27: 447 and compare 27: 1366. For evidence that he took lying to require the intention to deceive as late as the *Metaphysics of Morals*, see MM 6: 430. Kant also seems to imply that the polite lie 'your obedient servant' is not a lie because 'no one is deceived by it' (MM 6: 431). Politeness is 'a beautiful illusion resembling virtue' (MM 6:473). Unlike deception (*Betrug*), illusion (*Schein*) is compatible with true belief and so does not undermine rational choice (*Anthropology* 7: 149-50). A wig makes a bald man look hirsute even when there is common knowledge that he wears a wig. Declaring servitude makes Kant look obedient even to those who know he is not. The distinction between illusion and deception also explains Kant's remark, 'Poets are not liars except in panegyrics' (Reflection 618, 15: 266).

⁶ The English and German dictionaries of Kant's era do not require that a liar intend to deceive (Samuel Johnson (1775) and Johann Christoph Adelung (1774-86)). Almost all twentieth century philosophers require an intention to deceive. Bald-faced lies ('bare faced' in British idiom) have broken the consensus. A bald-faced lie made under oath might be made with the intention not to deceive anyone (Carson 2010, p. 21). The more obvious the perjury, the more reassured the witness-tampering mobster will be that his corrupted witness will continue to do his mendacious bidding.

⁷ For a compilation of Kant's jokes with helpful scholarly discussion, see Clewis (2020).

⁸ In Aquinas's terminology, falsehood is 'material falsehood' and believed falsehood is 'formal falsehood'. He overestimates the extent to which this distinction brings him into alignment with Augustine. Aquinas underestimates the extent to which sincere speech deviates from assertions believed true. Slips and malapropisms are pervasive because speech is mostly automatic and

speakers can rely on listeners to redact. Careful speech is a waste of time when sloppy speech is easily fixed on the fly. Absence of intentional editing into believed falsehood suffices for sincerity.

⁹ In the *Vigilantius* ethics lectures of 1793 Kant sketches the following division of the concept of *intentional deception*. Intentional deception, divides into two species: intentional deception by non-verbal means (as when one deceitfully gives the impression one is leaving on a journey by packing luggage) *versus* intentional deception by declaration (that is, by making a statement). The latter is both a *falsiloquium* and a lie in the ethical sense. A lie in the ethical sense divides into an action which is also a lie in the juridical sense (because another individual is injured), and a mere (deliberate) untruth with no injury to another, on the other. (27: 699-701). Clearly, according to this division, lies – whether juridical or merely ethical—form a species of deception. The view that lies in the ethical sense require ‘a second person whom one intends to deceive’ appears in the *Metaphysics of Morals* at 6: 430. Kant is especially concerned to argue—against what he sees as a traditional view to the contrary—that one need not injure any *particular* person in order to tell a lie in the ethical sense. Instead, he holds that one who tells a lie in the ethical sense injures humanity in general even if he does not injure any particular individual. The injury consists in the damage done to the institution of trust (6: 429, 6: 430).

¹⁰ Wood claims that for Kant a ‘declaration’ is a statement that occurs in a context in which ‘others are warranted or authorized in relying on the truthfulness of what is said’ (2011, p. 241). Wood, however, offers no evidence for this claim and we agree with Thomas Carson’s (2010, p. 75) that there is none. But even if Wood is correct, this would not imply that a declaration need be made with an intention to deceive.

¹¹ That characterization runs as follows ‘In the doctrine of right **an intentional untruth** is called a lie only if it violates another’s right; but in ethics, where no authorization is derived from harmlessness, it is clear of itself that *no intentional untruth in the expression of one’s thoughts can refuse this harsh name*’ (6: 429, emphasis added).

¹² In *The Conflict of the Faculties* (published in 1798, but compiling works written earlier) when reporting his notorious promise to King Friedrich Wilhelm II, Kant remarks that now in his 71st year he can’t help thinking he may have to answer ‘very soon’ to ‘a judge of the world who scrutinizes men’s hearts’ (7: 9-10).

¹³ Kant never explains why the throat grabber nonetheless persists with the inquiry. But by 1797, when he abandons the requirement that the liar intend to deceive, the anomaly is bypassed. Kant does not signal lane changes (Carson 2010, p. 78). His followers over-shoot and collide.

¹⁴ The London engraver John Spilsbury produced the first jig-saw puzzle in 1760. His ‘dissected maps’ of the world would have provided a more accurate analogy for Kant. Progress on a jig-saw puzzle is holistic and completeness is salient when the last piece snaps into place. In a 1770 letter to Johann Herbert Lambert, Kant predicts the *Critique* will be irrevocable and sticks to his prediction despite revisions about which particular doctrines are irrevocable (Kuehn 2001, pp. 231-234).

¹⁵ For a fuller account of these changes and their motivations, see Proops (2021, pp. 104-205).

¹⁶ For a detailed account of this argument, see Proops (2021, pp. 151-154).

¹⁷ *Real Progress*, 20: 308. There is no article before ‘unity’.

¹⁸ Confusingly, Kant does not use ‘simple’ in the second paralogism to mean ‘lacking parts’. Instead, he uses it—as well as ‘logically simple’—to mean ‘not plural’. To say of the soul (understood as ‘the thinking I’) that it is ‘logically simple’ in this sense is to say that it is not a group mind. This fact explains the otherwise baffling occurrence of ‘consequently’ in the third line of the remark we quote from B 407-8. No less confusingly, when Kant depicts the conclusion of the second paralogism in A as being that the soul is ‘one substance’ (A 352), he uses ‘substance’ non-

technically to mean ‘thing’ or ‘subject’ (See Proops (2021, p.140, fn 1—and for an explanation of the notion of ‘substance,’ that occurs in that note, see Proops (2021, p. 463).

¹⁹ Kant uses ‘ground of proof’ sometimes for an argument and sometimes for its premises. Either reading is acceptable in the present context.

²⁰ In therapeutic correspondence with Maria von Herbert, Doctor Kant characterizes the liar as *knowing* the falsehood of what he says. Was Kant lying about lying? Yes, if he over-simplified for persuasive effect. No, if he forgot his commitment to attributing lies solely on the basis of the speaker’s mental state (the point of his claim that lies track truthfulness rather than truth). Lies depend solely on the agent, not other agents (through their inferences) or the environment. Self-awareness of ignorance is compatible with lying. David Hume told some lies even if he was an extreme sceptic – and even if, indeed, nothing is known.

²¹ Neither can Kant (6: 430). Instead of concluding that lying to yourself is impossible, Kant at this stage concludes that these especially wicked lies must be possible in some way that eludes his understanding.

²² Chisholm and Feehan, 154; Collins’s notes of Kant ethics lectures from the period 1774-1777, 27: 447.

²³ For evidence that Kant regarded Voltaire as a paradigmatic indifferentist, see Proops (2021, pp. 265-270).

²⁴ This essay has benefitted from comments and suggestions by: Thomas Carson, Ed Holland, Jennifer Mensch, Allen Wood, and two anonymous referees for this journal.