

Lying: Knowledge or Belief?

Neri Marsili*

Forthcoming in *Philosophical Studies*

Abstract: A new definition of lying is gaining traction, according to which you lie only if you say what you know to be false. Drawing inspiration from “New Evil Demon” scenarios, I present a battery of counterexamples against this “Knowledge Account” of lying. Along the way, I comment upon the methodology of conceptual analysis, the moral implications of the Knowledge Account, and its ties with knowledge-first epistemology.

Keywords: Definition of Lying; Insincerity; Knowledge-first Epistemology; Ethics of Lying; Metaphilosophy; Conceptual Analysis

1. Lying and knowledge

He who cannot lie, doth not know what truth is
Also spoke Zarathustra

Suppose I tell you something that I believe to be false: for instance, that *it is raining in Athens*. If I believe (say, with good reasons) that it is *not* raining in Athens, but I don’t in fact *know* that it is not raining there, am I lying? More generally, is lying a matter of asserting what you *believe* to be false, or of asserting what you *know* to be false?

The traditional view is that the notion of lying is tied to belief, rather than knowledge. Philosophers typically accept that (A) and (B) are necessary¹ conditions for lying:

S lies only if:

(A) S asserts that p

(B) S believes that $\neg p$

Recently, however, it has been argued that (B) is too lenient. The suggestion is that lying should rather be tied to the notion of *knowledge*. To return to our initial example, I would be lying only if I *knew* that it’s false that it’s raining in Athens; *disbelieving* it is not enough for my utterance to be a lie. If this is correct, (B) should be replaced with (K):

(K) S knows that $\neg p$

Call the view that lying requires (K) instead of (B) the *Knowledge Account of Lying*. This view is only recently gaining traction, but it does have its precursors. To my knowledge, its earliest formulation is in Coignet’s *Politique discourses upon Truth and Lying* (1586), where lying is defined as a communicative act whereby “one speaketh of things vncertain, contrarie to that which one knoweth”². Note, however, that Coignet only requires that the speaker say something that *contradicts* what they know,

* Alma Mater Studiorum – Università di Bologna. Contact: nerimarsili@gmail.com

¹ There is disagreement about whether sufficiency requires that *S intends to deceive* the audience (Lackey 2013; Meibauer 2014; Mahon 2015; Fallis 2018; Dynel 2018; Stokke 2018; Krstić 2019; Wiegmann and Meibauer 2019, Marsili 2021, §2.3). This issue can be safely set aside for our purposes. For an approach to defining lying that relies on identifying *prototype elements* instead of *necessary and sufficient conditions*, see Coleman & Kay (1981).

² Some of Sartre’s remarks about bad faith point in a similar direction: “The essence of the lie implies in fact that the liar actually is in complete possession of the truth which he is hiding. A man does not lie about what he is ignorant of” (Sartre 1956, 48).

which is weaker than (K). For a formulation closer to (K) we have to wait until the XXI century. Saul (2000) is probably the first to claim, in passing³, that ‘lying requires knowingly speaking falsely’. Caminada (2009) endorses (K) more explicitly; like Saul, however, he doesn’t motivate his preference for (K) over (B). It’s only in the last decade that the Knowledge Account has been explicitly motivated and defended, in the works of Turri (2016, 34–36), Benton (2017), and Holguín (2019).

The question of whether lying requires knowledge or belief may seem philosophically marginal, but there is more at stake than the analysis of a concept, like lying, that is central to human communication. First, there are ethical implications. If (against traditional views) lying requires knowledge rather than belief, we would have to significantly rethink how we understand the moral blameworthiness of lying: whether one is responsible for having lied would depend not only on what one believes, but also on factors that are not (or that are only partially) under the speaker’s control, such as whether the proposition is actually false, whether the belief is Gettiered, and so forth. This isn’t a straightforward take on what makes lying blameworthy, and we will see that assessing the merits of this view requires some sophisticated reflection on how intentional action is assessed more generally.

Second, if the Knowledge Account turned out to be correct, this result would speak in favour of the knowledge-first paradigm in epistemology: the idea, gaining traction over the last two decades, that knowledge is a more fundamental concept than belief – an ‘unexplained explainer’ that can help us understand other key concepts that are central to epistemological theorising (Williamson 2000; McGlynn 2014), including lying. If the notion of lying is indeed best understood in terms of knowledge rather than belief, we would have new proof of the explanatory power of knowledge-first epistemology; if not, we may see this as a novel reason to question its ability to revolutionise our way of thinking about basic philosophical concepts.

I will come back to both points later in the paper. In the next section, I present and assesses the positive case for preferring (K) over (B). Section 3 shows that the Knowledge Account is subject to several counterexamples that are avoided by traditional definitions. Section 4 explores the moral implications of the Knowledge Account in detail. Section 5 introduces a last set of counterexamples, and section 6 considers the viability of defending the Knowledge Account by appealing to error theories. Section 7 concludes.

2. Reasons for preferring knowledge over belief

Why should we think that (K) is preferable to (B)? Benton, Holguín, and Turri defend the Knowledge Account by appealing to what we may call *coherence arguments*. Broadly, the idea is that we should endorse the Knowledge Account because it coheres with philosophical views that are independently supported by sound arguments and compelling empirical evidence.

First, it is noted that the Knowledge Account sits well with knowledge-first epistemology. A key contention of knowledge-first epistemology is that several concepts that are central to epistemological theorising are best explained in terms of knowledge, rather than (e.g.) belief (Williamson 2000, ch.1-2; McGlynn 2014, 16). If one is sympathetic to knowledge-first epistemology, one has a motivation to prefer (K) over (B). Second, the Knowledge Account has on its side its “elegant simplicity and pleasing symmetry” (Turri 2016, 34). Third, and perhaps most importantly, it is reinforced, indirectly, by the *philosophical* and *empirical* case in favour of the knowledge-norm of assertion⁴.

Let’s focus on this last point, since it needs unpacking. The idea is that the Knowledge Account is a natural way to define lying for those who accept the knowledge-norm of assertion: the view that *you are epistemically entitled to assert that p only if you know that p*. As Holguín (2019, 5352) puts it, “if we think there are knowledge-theoretic norms of assertion, we should expect the concept of lying to be intimately connected with the concept of knowing”.

Here’s a more explicit formulation of the idea (following Turri 2016, 34). Suppose that knowledge is the norm of assertion: you are entitled to assert that *p* only if you know that *p*. Whenever you assert a proposition that you don’t know, you violate the knowledge-rule. Suppose, further, that the term ‘lying’ denotes a special infraction of the norm of assertion: we use it to designate “cheating” at asserting. Arguably, to ‘cheat’ one must *knowingly* (that is, intentionally and

³ Saul (2012) has since moved on to defend a traditional view, preferring (B) over (K).

⁴ For a critical overview of the philosophical case for the knowledge-norm, McGlynn (2014, chap. 5); for the empirical case, Turri (2017).

successfully) break a rule. It follows that you lie when you knowingly break the rule of assertion – when you assert what you know to be false.

However, this argument only *seems* to establish the desired connection between (K) and ‘cheating at assertion’. If we accept its premises, we can only conclude that you lie when *you know that you don’t know* that *p*. This isn’t enough to prefer (K) over (B), because this requirement is weaker than both (K) and (B) – it can be met by a speaker who sincerely believes that *p* and doesn’t know that *p* is false⁵. Turri’s argument for deriving (K) from the knowledge-norm of assertion falls short of its intended goal⁶.

Suppose, however, that we find this set of coherence arguments persuasive. Even so, there are two important limitations that apply to them. First, coherence arguments offer at best *conditional* reasons to accept the Knowledge Account. This is a problem because both the knowledge-norm of assertion and the knowledge-first paradigm have been subject to substantial criticisms⁷, and neither can be considered a consensus view. If these criticisms are sound, and knowledge isn’t the norm of assertion (nor is it more fundamental than belief), we lack a positive case to prefer (K) over (B) – because the antecedent of our conditional argument is false. The result is that the positive case in favour of (K) is pretty weak: only *if* a certain set of views is correct, we have a positive reason to prefer (K) over (B) – and since the correctness of these views is disputed, it is equally disputed that we enjoy this positive reason to prefer (K) over (B).

Second, even if the antecedent of the conditional argument is true (and we have a positive reason to accept the Knowledge Account), the problem remains that coherence arguments alone cannot provide us with *conclusive* reasons to accept the Knowledge Account. This is because philosophers tend to agree that *intensional accuracy* (*i.e.* whether a definition systematically tracks our intuitions) is the primary desideratum for a definition of lying (Fallis 2009, §1; cf. Gupta 2015, §1.4), and that this desideratum must be met for a definition to be informative in our moral theorising about lying (Marsili 2019). Clearly, we cannot determine that a definition is intensionally accurate on the basis of coherence arguments alone. If it turns out that replacing (B) with (K) leads to predictions that systematically clash with our semantic and moral intuitions about lying, insisting that (K) coheres with existing philosophical views will do little to prove that the Knowledge Account is a correct definition of lying (rather than, say, a useful theoretical construct).

Summing up, the positive case for the Knowledge account is somewhat weak. We cannot easily derive (K) from the knowledge-norm of assertion. And coherence arguments can only provide *conditional, non-conclusive* reasons to accept the Knowledge Account. Of course, this only means that we have little reason to *endorse* the view, not that we should *reject* it. But we will now see that there is also a solid negative case *against* the Knowledge Account.

3. The New Evil Liar

Here’s a first way to challenge the idea that we can only lie about what we know to be false. Consider the following scenario, in the spirit of New Evil Demon scenarios (Littlejohn 2009):

NEW EVIL LIAR

Pinocchio saw that Lucignolo stole an apple from his classmate Berlicche. Based on this experience, he forms the correct belief that Lucignolo stole Berlicche’s apple. When Pinocchio notices that Berlicche is looking for his apple, he tells him:

- (1) Eugenio stole the apple from your desk

Pinocchio has a twin, Malocchio, who is the unfortunate inhabitant of an evil demon world. Based on experiences indistinguishable from those of Pinocchio, Malocchio forms the belief

⁵ For example, suppose that I believe *p* on a hunch, for no good reason. I *know that I don’t know* that *p*. But if I assert *p*, I would meet neither (K) nor (B).

⁶ Benton (2018) proposes a similar derivation, but acknowledges the argument’s limitations. Holguín (2019) derives (K) from the knowledge norm and the stronger premise that one lies when “one asserts the negation of what one ought to have asserted”. This gets us to (K), but at the price of quasi-circularity: Holguín’s extra premise (for which no argument is provided) is just (K) in disguise.

⁷ See, e.g., Pagin (2016) and Marsili (2018b) for criticisms of the former, and McGlynn (2014) and Gerken (2017) for criticisms of the latter. A criticism of the empirical case in favour of the norm of assertion is in Marsili and Wiegmann (2021).

that Lucignolo stole an apple from Berlicche. Just like Pinocchio, Malocchio tells Berlicche that (1).

Intuitively, both Pinocchio and Malocchio have lied to their respective classmates. However, while Pinocchio has asserted something that he *knows* to be false, Malocchio cannot know that what he said is false, because he inhabits an evil demon world. According to the Knowledge Account, Pinocchio lied, but Malocchio didn't.

This seems intuitively wrong⁸. Pinocchio's and Malocchio's actions were motivated by the same evidence and doxastic states, and by the same malicious intentions: intuitively, if one has lied, also the other one has. The Knowledge Account is committed to denying this. Since the two scenarios are identical from the speaker's perspective, the Knowledge Account also entails that whether a speaker is lying may depend on factors that are beyond the speaker's control. This is another controversial commitment, which has problematic moral implications. If (as the Knowledge Account has it) only Pinocchio lied, his behaviour is morally criticisable in a way in which Malocchio's behaviour is not: both would be guilty of *attempted deception*, but only Pinocchio would be guilty of *lying*. Intuitively, however, Pinocchio and Malocchio are equally blameworthy: after all, they acted in the same way, with the same intentions and consequences, and for the same reasons⁹. This places an additional burden on the Knowledge Account: its proponents need to offer a plausible story for introducing a morally loaded distinction where there seems to be none.

Admittedly, New Evil Demon scenarios are no everyday situations, and as such it may be maintained that they have limited evidential weight. In response, we could consider a more realistic NED scenario:

NEW EVIL HOLE

Pinocchio owes 15 liras to Lucignolo, who asks for his money back. Pinocchio knows that he put 20 liras in a hidden pocket of his schoolbag just before leaving home. Nonetheless, he tells Lucignolo:

(2) I don't have 15 liras on me right now

Pinocchio has a twin, Bucocchio, who lives in a different town. Like Pinocchio, Bucocchio owes Lucignolo's twin 15 liras, has 20 liras hidden in his pocket, and has told Lucignolo's twin that (2).

But Bucocchio is unaware that his pocket has developed a hole since he last used it. Some of Bucocchio's money has fallen out of his pocket since he left home.

⁸ Holguín (2019) discusses a similar case, and attempts to explain away this prediction of the Knowledge Account. I will discuss his arguments in sections 4 and 6. A referee suggests that, since the internal processes underlying Malocchio's belief are isomorphic to Pinocchio's, it may be argued that *both meet the standard for knowing* and, thus, that *both are lying* in uttering (1). In reply, even if we conceded that Malocchio knows that (1) is false, we could not infer from this that he is lying: this inference would treat (K) as a premise, but our aim here is precisely to establish whether (K) holds. Furthermore, this strategy is unlikely to generalise. I will soon discuss NED scenarios where the twin lacks different requirements for knowledge (*safety, truth, justification, confidence*, etc.). To argue that the twin can meet the standard for knowing in each case would force us to revise our concept of knowledge beyond recognition, and (K) would end up being a very different thesis (one much closer to (B)). Additionally, and crucially, in §4 I will consider a case that does *not* involve internally isomorphic twins. For more on revisionist replies, cf. §6.

⁹ Can we avoid this prediction simply by assuming that lying is just as morally bad as misleading? Endorsing this view may help *narrow* the gap between the predictions of the Knowledge Account and our moral intuitions and practices, but it won't *fill* the gap altogether. Even authors who defend the thesis that misleading is just as bad as lying acknowledge (i) that lying and misleading are different *kinds* of infractions, and (ii) that, as a matter of fact, we associate different moral stigmas and different social sanctions to each (cf. Saul 2012, esp. §4.4). The Knowledge Account's verdict that Pinocchio and Malocchio have committed *different* kinds of infraction is therefore problematic *regardless* of one's take about the moral difference between lying and misleading, since it clashes with both (i) and (ii). For more on the moral implications of (K), see §4.

Also in this scenario, both Pinocchio and Bucocchio intuitively lied. However, Bucocchio doesn't know that what he said is false. The Knowledge Account rules that Pinocchio has lied and Bucocchio hasn't. But this doesn't seem right: after all, this scenario is structurally similar to NEW EVIL LIAR, except that it involves local deception (Bucocchio is mistaken about his finances), rather than universal deception (Malocchio knows nothing about the external world). Furthermore, this classification licenses morally loaded judgments that seem equally doubtful: while Pinocchio would be guilty of lying, Bucocchio could at most be criticized for attempting to deceive.

Note that the scenario does not specify whether Bucocchio actually has fifteen liras. If more than 5 liras fell out of Bucocchio's pocket (TRUE-VERSION), Bucocchio doesn't know that (2) is false because (2) is true. If less than 5 liras fell out of Bucocchio's pocket (GETTIER-VERSION), Bucocchio doesn't know that (2) is true because his belief is unsafe and based on misleading evidence.

I left both options open because not all readers will find the TRUE-VERSION to be a straightforward case of lying. While most philosophers¹⁰ and laypeople¹¹ would agree that Bucocchio is lying even if it turns out that more than 5 liras fell out of his pocket, a few readers may not share this intuition. But this is no reason to find the GETTIER-VERSION any less convincing. In the GETTIER-VERSION there is even *less* difference between what Pinocchio and Bucocchio have done (both have said something false), making it even less plausible to claim that one is lying while the other isn't¹². The NEW EVIL HOLE scenario is effective because it offers a counterexample for every taste: most people will find both versions convincing, and the few who are unpersuaded by TRUE-VERSION should still be persuaded by its GETTIER-VERSION¹³.

4. Blameworthiness and internal requirements for knowledge

I have argued the Knowledge Account makes predictions that are hard to reconcile with (i) our intuitions about what lying is, and (ii) our intuitions about the *moral* difference between each pair of examples. Here I would like to consider whether something can be said in response to the second criticism.

The accusation is that the Knowledge Account introduces a morally loaded distinction where there seems to be none. For instance, many will intuit that whether you are criticisable for having lied should not depend on whether your belief in what you said is Gettiered or not. Controversially, the Knowledge Account rules otherwise. Can something be said to explain away this prediction?

Here's a way to respond on behalf of the Knowledge Account: in both the "good cases" (where the speaker has knowledge) and the "bad cases" (where the speaker lacks it) of every scenario considered so far, the speaker *believes* that they are lying – that is, *they believe that they know* that what they said is false. For instance, in NEW EVIL LIAR (where Malocchio's belief is Gettiered, but not Pinocchio's), both Pinocchio and Malocchio *believe* that they know that their assertion is false. If we have the intuition that Pinocchio and Malocchio are equally blameworthy, it may be suggested, it is not because they both lied, but rather because *they both believe that they lied*.

Holguín (2019, 5366) has hinted at the possibility of a reply along these lines. He mentions that the moral implications of the Knowledge Account may be reconciled with our intuitions by appealing to a principle like (P):

(P) If as a result of X-ing A is (or would be) subject to normative scrutiny, then for any other person B that is in the relevant respects an internal duplicate of A, B is subject to *some* amount of normative scrutiny too¹⁴

¹⁰ Augustine (DM), Aquinas (ST), Kant (1797), Leonard (1959), Isenberg (1964), Mannison (1969), Chisholm and Feehan (1977), Kupfer (1982), Adler (1997), Williams (2002), Mahon (2015), Horn (2017), Wiegmann (forthcoming). For the opposite view, Sartre (1956); for intermediate views, Siegler (1966), Carson (2010), and Saul (2012).

¹¹ Turri and Turri (2015) found mixed evidence about people's intuitions: participants would respond in radically different ways, depending on how the question was asked. A subsequent study by Wiegmann, Samland, and Waldmann (2016) has shown that people consistently judge true lies to be genuine lies, and that the mixed findings by Turri and Turri were the by-product of a flawed questioning method (cf. Marsili 2016, 299, Wiegmann, Rutschmann, and Willemsen 2017, §3, Wiegmann and Meibauer 2019). In light of these criticisms and further empirical research, Turri and Turri (2019) retreated to the weaker (and much less controversial, cf. Coleman and Kay 1981) claim that only *prototypical* lying requires falsity.

¹² We may even construct the scenario so that *no* liras fell off the pocket: Bucocchio's belief would be lucky and unsafe anyway, preventing it from being knowledge (Sosa 1999; Pritchard 2005).

¹³ Of course, those who intuit that Bucocchio is not lying in the TRUE VERSION will be dissatisfied with both (K) and (B), since (K) makes incorrect predictions in the GETTIER VERSION, and (B) in the TRUE VERSION. To accommodate the intuition that lies must be false while avoiding the pitfalls of (K), a further condition (e.g. T: the proposition is false) could be added to the definition.

¹⁴ Holguín (2019, 5366). Italics are mine.

If (P) is correct, we would have an alternative explanation for our intuition that Pinocchio and Malocchio are equally blameworthy: the reason is not that they both lied, but rather that their actions were the result of internally identical processes: both intended to bring about the same consequences, and both thought that they were committing the same morally reprehensible act.

This explanation may seem promising, but there are two important difficulties that prevent it from being fully persuasive. The first is that (P) is too weak. It only tells us that an internal duplicate is subject to *some* normative scrutiny, not that internal twins are subject to *equal* normative scrutiny. This doesn't help much: recall that the intuition that we need to explain away is that Pinocchio and Malocchio seem to have committed the same kind of infraction, and therefore that they are equally morally blameworthy, and liable to the same sort of criticism.

To address this worry, we could amend (P), and replace “some” with “the same”. This modified version of (P) – call it (P*) – would correctly predict that the twins in our examples are equally blameworthy. However, (P*) is a much more controversial principle than (P), and as such it stands in need of proper justification. To illustrate its contentiousness with an example: according to (P*), attempted murder should be subject to the *same* moral scrutiny as successful murder. While there is a sense in which both *attempted* and *actual* murder reflect equally badly on the moral character of an agent, it is at least disputable that murder and attempted murder are equally morally reprehensible, or liable to the same sort of criticism¹⁵. While (P) is too weak, there is a risk that (P*) may be too strong.

Suppose that we accept (P*) despite these worries. We would then face a second, more pressing problem: namely, that counterexamples to the Knowledge Account need not involve internal duplicates. To illustrate, consider the following scenario:

EVIL AUNTIE

Grandpa Joe recently died. Pinocchio, who's very superstitious, organises a seance with a medium to try to communicate with Grandpa Joe one last time. During the seance, it's revealed that Grandpa Joe left all his money to a woman. On this basis, Pinocchio forms a firm belief that Grandpa Joe left all his money to Auntie Sue. Pinocchio was in fact meant to see a notary today – to discuss, amongst other things, Grandpa Joe's will. Believing that Pinocchio has already met with the notary, Auntie Sue calls Pinocchio, asking if there are any news about the will. Pinocchio is mad at Auntie Sue. Out of pure spite, he replies:

(3) It turns out that Grandpa Joe left all his money to charity

Auntocchio inhabits a twin world where everything holds the same. The only difference is that Auntie Sue calls him *after* Auntocchio saw the notary, and that the notary has confirmed that Grandpa Joe left everything to Auntie Sue.

Both Pinocchio and Auntocchio have intuitively lied to Auntie Sue, and they are both criticisable for having lied. However, while Auntocchio *knows* that what he said is false, Pinocchio only *believes* that what he said is false – he cannot know it, for his belief is based on *unsafe* evidence¹⁶. The Knowledge Account incorrectly predicts that only Auntocchio lied, and therefore that only Auntocchio is blameworthy for having lied.

The example is interesting for two reasons. First, because it provides a novel kind of counterexample to the Knowledge Account. In this scenario, Pinocchio's belief is not merely *unlucky*. It is based on blatantly poor evidence. Furthermore, Pinocchio and Auntocchio are not internal duplicates: their perspective on the external world (and the evidence available to them) differs significantly. While Auntocchio has access to *good, safe reasons* (the notary's testimony) to believe that what he said is false, Pinocchio has only access to *poor, unsafe reasons* (the seance) in support of the same belief.

¹⁵ On the moral and legal difference between successful and attempted crimes, see e.g. Davis (1986). For an overview on moral luck more generally, see Nelkin (2021). Note that I am not here denying that (P*) is defensible; I am merely pointing out that (P*) is a costly commitment, since (P*) takes a rather strong stance on philosophically controversial issues.

¹⁶ I am assuming that both Pinocchio and Auntocchio inhabit a world (like ours) where it is impossible to communicate with dead people: if Pinocchio is right about Grandpa's will, this is due to luck, not to sound reasoning.

Second, given that Auntocchio and Pinocchio are not internal duplicates, we cannot appeal to (P) or (P*) to explain away the intuition that Pinocchio and Auntocchio are guilty of the same kind of infraction. (P) and (P*) have simply nothing to say about cases that don't involve internal duplicates: our moral intuitions about this scenario cannot be reconciled with the Knowledge Account by appealing to these principles.

Recapitulating, the worry outlined at the beginning of this section seems hard to address. The Knowledge Account does face an explanatory burden that the Belief Account does not face: only the former view needs to be complemented with some principle that explains away our moral intuitions about the cases in which the speaker satisfies (B) but not (K). A principle like (P*) isn't likely to help: beyond being difficult to defend, it is useless for scenarios, like EVIL AUNTIE, that do not rely on internal duplicates. It seems difficult, then, to reconcile the Knowledge Account with our moral intuitions¹⁷. Judging by moral implications alone, the Belief Account fares better than its knowledge-based counterpart.

5. Graded-belief lies

Knowledge is incompatible with uncertainty: you know a proposition only if you outright *believe* that proposition¹⁸. But is outright belief required for lying? It has been argued that this requirement would be too strong (Carson 2006, 2010; Whyte 2013; Marsili 2014, 2018a, 2019; Krauss 2017; Trpin, Dobrosovetsnova, and Sebastian 2020; cf. Isenberg 1964; Benton 2018). If (K) requires outright belief, it incorrectly rules out *graded-belief lies*. Consider the following scenario:

EVIL CURRY

Currocchio is having friends for dinner. A guest asks him if the curry contains nutmeg, explaining that she has a mild intolerance. Currocchio is quite confident that he did put nutmeg in the curry, although he isn't sure: he cooked the curry the previous day, and added lots of random spices. Since he can't be bothered preparing something new, he simply replies:

(4) I didn't put any nutmeg in the curry

Currocchio plainly stated something that he takes to be likely false. In doing so, intuitively he lied. There is consensus that this is the right verdict: while philosophers disagree about which definition best handles *graded-belief lies* like (4), there is agreement that graded-belief lies are intuitively lies¹⁹. A definition that rules that Currocchio has not lied is perhaps defensible, but in need of a strong independent argument to support its verdict. Once again, (K) is in a weak position to explain our intuitions²⁰.

6. Explaining counterexamples away

We have seen that the Knowledge Account clashes, sometimes even sharply, with our pre-theoretical intuitions about what lying is. If so, how can its proponents maintain that (K) is preferable to (B), the traditional view? Among the defenders of (K), only Holguín (2019) has attempted to provide an answer.

Holguín's strategy is to question the reliability of our intuitions about each example. For instance, in response to counterexamples based on graded-belief lies, he suggests (i) that lying and bullshitting

¹⁷ Earlier I considered the idea that *believing* that you committed a moral infraction is sufficient for being blameworthy for that action. One may wonder whether it is rather this principle that can explain why the twins in each scenario are equally blameworthy. Leaving aside the fact that this alternative principle is less plausible than (P*), I shall note that it is easy to amend EVIL AUNTIE so that this requirement is not satisfied: we simply need to specify *ex hypothesi* that Pinocchio doesn't believe that he knows that (3) is false. For instance, Pinocchio may be aware that mediums are unreliable sources, but form his belief (perhaps irrationally) *despite* being aware of this. More generally, this alternative principle will not work whenever the speaker holds a belief despite being aware that that belief is based on poor or unreliable evidence.

¹⁸ This is a requirement that knowledge-firsters accept. See, for instance, Williamson (2000, §4).

¹⁹ For a detailed defence, see Krauss (2017) and Marsili (2014, 2018a, 2019), who provide competing explanations for this datum. See Holguín (2019, 5361-4) and §6 of this paper for an argument on whether we should *doubt* the intuition that graded-belief lies are genuine lies.

²⁰ (K) is not alone: since (B) too requires outright belief, also (B) incorrectly predicts that Currocchio isn't lying. Luckily, there are alternative views that avoid this unpalatable prediction, on which I will come back in the next section.

are mutually exclusive categories, and (ii) that “S bullshits iff S knows that S does not know whether p”. These stipulations entail that *graded-belief lies* are not genuine lies, but rather ‘bullshit’, so that our intuitions are mistaken whenever we classify a statement like (4) as a lie. If our intuitions about graded-belief lies like (4) are mistaken, such ‘counterexamples’ pose no real challenge to (K). Similar error theories are introduced to argue that our intuitions about Gettiered lies (like NEW EVIL LIAR) and ‘true lies’ (like the TRUE-VERSION of NEW EVIL HOLE) are systematically unreliable. Generalising, the idea is that (K) can be defended by introducing error-theories that explain away our intuitions about each counterexample. Call a response along these lines an ‘error-theoretic response’.

Considering the merits of each error theory introduced by Holguín in defense of (K) would require a long digression, but I worry that such a detour would be both unnecessary and ineffective. It would be *unnecessary* because this paper has already introduced counterexamples to (K), such as the EVIL AUNTIE case, that evade Holguín’s defense²¹. Even if Holguín’s error-theories were all successful in explaining away their target counterexamples, we would still have that (K) makes incorrect predictions about all those cases in which the speaker meets (B), but has undisputably poor evidence in support of their belief (like in EVIL AUNTIE).

Discussing Holguín’s proposed cocktail of error theories would also be an *ineffective* way to attack the Knowledge Account, because it would not rule out the possibility that (K) could be defended by appeal to some *other* combination of error theories. If we could prove that error-theoretic responses are *in principle* improper, by contrast, we would have a much stronger argument against the Knowledge Account. In the rest of this section, I will attempt to offer a general reply along these lines.

Let me start by noting an apparent tension between the methodology of conceptual analysis and the idea that we can explain away counterexamples by appealing to error theories about our intuitions. We have seen (§2) that the main goal of authors who attempt to define lying is to develop an *analysis* of this concept, which requires tracking ordinary language intuitions about whether a given utterance is a lie. Within this kind of inquiry, intuitions are the raw data against which competing definitions are tested²². Postulating error theories about our intuitions amounts to excluding some of the data (typically, and problematically, data that would falsify one’s favourite view). If postulating error theories about our intuitions is acceptable at all when engaging in conceptual analysis (as opposed to conceptual engineering), then, it would seem that it is when it is done (i) parsimoniously, (ii) with sound motivations, and (iii) in absence of more economic alternatives. The envisaged strategy for defending (K) doesn’t meet any of these conditions.

If we are to rescue (K), error-theories cannot be postulated sparingly. This point is illustrated by Holguín’s (2019) own defence, which has to²³ harvest, one after another, several revisionist theories about our intuitions about knowledge and lying (e.g. about the difference between *lies and true lies*, the difference between *lying and bullshit*, the possibility of *unjustified knowledge*). Unsurprisingly, being generous with such revisionary assumptions generates difficulties. First, by simple probability calculus, each ancillary assumption automatically decreases the probability that the conjunction of views is correct. Second, each ancillary assumption brings us closer to providing a revisionary account of lying, and thus farther from our aim (a descriptive analysis of the concept). Third, the motivation to endorse any specific combination of revisionary assumptions (rather than another) cannot be that that specific combination squares our intuitions with the Knowledge Account, on pains of circularity. But it isn’t clear what other reasons we might have to adopt such a specific collection of revisionary views. The charge of ad-hocness isn’t inescapable here, but it lurks around the corner for the defender of the Knowledge Account – and existing defences, like Holguín’s, have done little to counter it.

²¹ Arguably, also EVIL CURRY is a counterexample for which no error theory has been provided yet. Holguín only discusses graded belief in relation to lottery cases, where the speaker’s belief is formed on the basis of statistical evidence. EVIL CURRY relies on a different kind of uncertainty (compromised memories), and a plausible case can be made that Holguín’s argument is less effective when applied to this sort of case. For a further counterexample to (K) for which Holguín provides no error theory, cf. Marsili’s (2016) empirical study on believed-true lying promises.

²² This is something on which there is solid agreement within the literature, and this very assumption is what motivates the empirical research that has been conducted about laypeople’s concept of lying. For discussion, see Carson (2006, 285; 2010, 33-7), Fallis (2009, 32), Arico & Fallis (2013, §3), Weissman & Terkourafi (2018, §4), Marsili (2019, §2), Reins & Wiegmann (2021, §2.3).

²³ The emphasis is on ‘has to’. (K) imposes several requirements on lying: that the lie be *false*, that the speaker’s belief be *safe, justified, not graded*, etc. Since each requirement is challenged by independent counterexamples, equally independent error-theories have to be introduced to explain them all away.

Perhaps these criticisms would be less pressing if we lacked more intuitive accounts of what lying is. But (and this is the crucial point) this isn't the case. There are definitions that deliver the right classifications at a much lower cost – that is, without invoking an array of exotic error theories. The Outright Belief account (B) avoids all the counterexamples discussed in section 3 and 4 (NEW EVIL LIAR, NEW EVIL HOLE, and EVIL AUNTIE). And if one is moved by counterexamples based on graded-belief lies (EVIL CURRY), more fine-grained accounts are available, such as Carson's (2006, 2010) *Weak Insincerity Condition*, or Marsili's (2014, 2019) *Comparative Insincerity Condition*²⁴:

(WIC) The speaker does not believe the proposition to be true

(CIC) The speaker takes herself to be more confident in the falsity of p than in its truth

These accounts can deal with all the cases discussed so far, without postulating any error theories about our intuitions. They avoid counterexamples inspired by 'New Evil Demon' scenarios, because they do not require knowledge. And they accommodate graded-belief lies, because they do not require outright belief²⁵.

In short, the contention here is that any putative error-theoretic defence of the Knowledge Account will not meet any of the above identified conditions (i-iii). If the aim is to provide an analysis of the concept of lying (as opposed to a revisionist account), views based on belief (like B) or credence (like WIC or CIC) should be preferred, since they track our intuitions far more reliably than the Knowledge Account.

7. Belief rather than knowledge

The Knowledge Account of lying is an appealing view. It sits well with influential philosophical theories, such as the knowledge-rule of assertion and the knowledge-first paradigm in epistemology, and has on its side its "elegant simplicity and pleasing symmetry". But a good definition must also be correct – that is, intensionally accurate. We have seen that Knowledge Account fails in this respect: it incorrectly excludes some straightforward cases of lying, and has counterintuitive moral implications. While it may be useful as an artificial theoretical construct, it isn't helpful as a descriptive analysis of what lying is.

We can finally answer our initial question: is lying a matter of asserting what you *believe* or *know* to be false? Perhaps surprisingly, we are better off sticking with traditional views: lying is a matter of saying what you disbelieve, rather than what you know to be false. Similarly, the distinctive moral blameworthiness associated with lying is intuitively a matter of internal doxastic states, and seems to be independent of any further, non-doxastic conditions for knowledge (such as whether the belief is correct, whether the reasons supporting it are safe, reliable, etc.).

These results, in turn, put some pressure on the ambitions of knowledge-first epistemology. The traditional, belief-first analyses of the concept of lying turned out to be more descriptively accurate than their knowledge-first alternative. While this is certainly a minor defeat for knowledge-first epistemology, this explanatory failure does add up to a growing list of concepts that knowledge-firsters struggle to handle within their framework²⁶. As such, this paper has also contributed to make a step (although a small one) towards reassessing the extent to which knowledge can be the game-changing "unexplained explainer" that knowledge-firsters want it to be²⁷.

²⁴ Also Krauss (2017) proposes a definition that accounts for graded beliefs, but it is known to be vulnerable to counterexamples (Benton 2018; Marsili 2019).

²⁵ As I mentioned in footnotes 11 and 13, a minority of laypeople and philosophers have the intuition that an assertion is a lie only if it is false, and this intuition can be accommodated by adding the further requirement that the proposition must be false.

²⁶ A growing body of literature, for instance, questions the possibility of analysing factive mental states (like seeing, perceiving or remembering) in terms of knowledge (McDowell 2002; Bernecker 2009; Pritchard 2011; 2013; Turri 2010; Schroeder manuscript; Cozzo 2010). For a broader overview of the obstacles encountered by the reductive ambitions of knowledge-first epistemology, see McGlynn (2014) and Gerken (2017).

²⁷ The author would like to thank Vladimir Krstić, Michele Palmira, Giorgio Volpe and Alex Wiegmann for their helpful feedback on earlier versions of this draft.

References

- Augustine. DM. *De Mendacio Liber Unus*. Accessible at Wikisource, URL: https://la.wikisource.org/wiki/De_mendacio
- Aquinas. ST. *Summa Theologiae*. Accessible at Wikisource, URL: https://la.wikisource.org/wiki/Summa_Theologiae
- Adler, Jonathan E. 1997. 'Lying, Deceiving, or Falsely Implicating'. *Journal of Philosophy* 94 (9): 435–52.
- Arico, Adam J., and Don Fallis. 2013. 'Lies, Damned Lies, and Statistics: An Empirical Investigation of the Concept of Lying'. *Philosophical Psychology* 26 (6): 790–816. <https://doi.org/10.1080/09515089.2012.725977>.
- Benton, Matthew Aaron. 2017. 'Lying, Belief, and Knowledge'. In *The Oxford Handbook of Lying*. <https://doi.org/10.1093/oxfordhb/9780198736578.013.9>.
- Benton, Matthew Aaron. 2018. 'Lying, Accuracy and Credence'. *Analysis* 78 (2): 195–98. <https://doi.org/10.1093/analys/anx132>.
- Bernecker, Sven. 2009. *Memory: A Philosophical Study*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199577569.001.0001>.
- Caminada, Martin. 2009. 'Truth, Lies and Bullshit; Distinguishing Classes of Dishonesty'. In *Proceedings of the 2009 IJCAI Workshop on Social Simulation*.
- Carson, Thomas L. 2006. 'The Definition of Lying'. *Noûs* 2: 284–306.
- Carson, Thomas L. 2010. *Lying and Deception*. Oxford: Oxford University Press.
- Chisholm, Roderick M., and Thomas D Feehan. 1977. 'The Intent to Deceive'. *Journal of Philosophy* 74 (3): 143–59.
- Coignet, Martyn. 1586. *Politique Discourses upon Trueth and Lying an Instruction to Princes to Keepe Their Faith and Promise: Containing the Summe of Christian and Morall Philosophie, and the Duetie of a Good Man in Sundrie Politique Discourses Vpon the Trueth and Lying*.
- Coleman, Linda, and Paul Kay. 1981. 'Prototype Semantics: The English Word Lie'. *Language* 57 (1): 26. <https://doi.org/10.2307/414285>.
- Cozzo, Cesare. 2010. *Is Knowledge the Most General Factive Stative Attitude?* In Carlo Cellucci, Emiliano Ippoliti & Emily Grosholz (eds.), *Logic and Knowledge*. Cambridge Scholars Press, 84–88.
- Davis, Michael. 1986. 'Why Attempts Deserve Less Punishment than Complete Crimes'. *Law and Philosophy* 5 (1): 1–32. <https://doi.org/10.2307/3504711>.
- Dynel, Marta. 2018. *Irony, Deception and Humour. Seeking the Truth about Overt and Covert Untruthfulness*. Berlin, Boston: De Gruyter, Mouton. <https://doi.org/10.1515/9781501507922>.
- Fallis, Don. 2009. 'What Is Lying?' *Journal of Philosophy* 106 (1): 29–56.
- Fallis, Don. 2018. 'What Is Deceptive Lying?' In *Lying: Language, Knowledge, Ethics, Politics*. Oxford University Press. <http://dx.doi.org/10.2139/ssrn.1702023>.
- Gerken, Mikkel. 2017. 'Against Knowledge-First Epistemology'. In *Knowledge-First Approaches in Epistemology and Mind*, edited by Adam Carter, Emma Gordon, and Benjamin Jarvis, 46–71. Oxford: Oxford University Press.
- Gupta, Anil. 2015. 'Definitions'. In *Stanford Encyclopedia of Philosophy* (Summer 2015 edition). <https://plato.stanford.edu/archives/sum2015/entries/definitions/>
- Holguín, Ben. 2019. 'Lying and Knowing'. *Synthese*, 198, 5351–5371. <https://doi.org/10.1007/s11229-019-02407-2>.
- Horn, Laurence R. 2017. 'Telling It Slant: Toward a Taxonomy of Deception'. In *The Pragmatic Turn in Law: Inference and Interpretation in Legal Discourse*, 1–27.
- Isenberg, Arnold. 1964. 'Deontology and the Ethics of Lying'. *Philosophy and Phenomenological Research* 24 (4): 463–80.
- Kant, Immanuel. 1797. *On a Supposed Right to Lie Because of Philanthropic Concerns*, translation of "Über ein vermeintes Recht aus Menschenliebe zu lügen", *Berlinische Blätter* 10, 301–314.
- Krauss, Sam Fox. 2017. 'Lying, Risk and Accuracy'. *Analysis* 73: 651–59. <https://doi.org/10.1093/analys/anx105>.
- Krstić, Vladimir. 2019. 'Can You Lie Without Intending to Deceive?' *Pacific Philosophical Quarterly* 100 (2): 642–60. <https://doi.org/10.1111/papq.12241>.
- Kupfer, Joseph. 1982. 'The Moral Presumption against Lying'. *The Review of Metaphysics* 36 (1): 103–26.
- Lackey, Jennifer. 2013. 'Lies and Deception: An Unhappy Divorce'. *Analysis* 73 (2): 236–48. <https://doi.org/10.1093/analys/ant006>.
- Leonard, HS. 1959. 'Interrogatives, Imperatives, Truth, Falsity and Lies'. *Philosophy of Science* 26 (3): 172–86.
- Lindley, T Foster. 1971. 'Lying and Falsity'. *Australasian Journal of Philosophy*, 49 (2): 37–41.
- Littlejohn, Clayton. 2009. 'The New Evil Demon Problem'. In *Internet Encyclopedia of Philosophy*.
- Mahon, James Edwin. 2015. 'The Definition of Lying and Deception'. In *Stanford Encyclopedia of Philosophy* (Fall 2015 edition). <https://plato.stanford.edu/archives/fall2015/entries/lying-definition/>
- Mannison, D.S. 1969. 'Lying and Lies'. *Australasian Journal of Philosophy* 2 (47): 132–44.
- Marsili, Neri. 2014. 'Lying as a Scalar Phenomenon'. In *Certainty-Uncertainty – and the Attitudinal Space in Between*, edited by Sibilla Cantarini, Werner Abraham, and Elisabeth Leiss, 153–73. Amsterdam: John Benjamins Publishing Company. <https://doi.org/10.1075/slcs.165.09mar>.

- Marsili, Neri. 2018a. 'Lying and Certainty'. In *The Oxford Handbook of Lying*, edited by Jörg Meibauer, 169–82. Oxford: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780198736578.013.12>.
- Marsili, Neri. 2018b. 'Truth and Assertion: Rules versus Aims'. *Analysis* 78 (4): 638–48. <https://doi.org/10.1093/analys/any008>.
- Marsili, Neri. 2019. 'Immoral Lies and Partial Beliefs'. *Inquiry*. <https://doi.org/10.1080/0020174X.2019.1667865>.
- Marsili, Neri. 2020. 'Lying, Speech Acts, and Commitment'. *Synthese*. <https://doi.org/10.1007/s11229-020-02933-4>.
- Marsili, Neri, and Alex Wiegmann. 2021. 'Should I Say That? An Experimental Investigation of the Norm of Assertion'. *Cognition*, 212, <https://doi.org/10.1016/j.cognition.2021.104657>.
- McDowell, John. 2002. 'Responses'. In *Reading McDowell: On Mind and World*, edited by Nicholas H. Smith, 269–305.
- McGlynn, Aidan. 2014. *Knowledge First?* Palgrave Macmillian.
- Meibauer, Jörg. 2014. *Lying at the Semantics-Pragmatics Interface*. *Lying at the Semantics-Pragmatics Interface*. Berlin, Boston: De Gruyter. <https://doi.org/10.1515/9781614510840>.
- Nelkin, Dana K. 2021. 'Moral Luck'. In *The Stanford Encyclopedia of Philosophy* (Summer 2021 edition). <https://plato.stanford.edu/archives/sum2021/entries/moral-luck/>.
- Pagin, Peter. 2016. 'Problems with Norms of Assertion'. *Philosophy and Phenomenological Research* 93 (1): 178–207. <https://doi.org/10.1111/phpr.12209>.
- Pritchard, Duncan. 2005. *Epistemic Luck*. Oxford: Clarendon Press.
- Pritchard, Duncan. 2011. 'Epistemological Disjunctivism and the Basis Problem'. *Nous-Supplement: Philosophical Issues* 21 (1): 434–55. <https://doi.org/10.1111/j.1533-6077.2011.00210.x>.
- Pritchard, Duncan. 2013. *Epistemological Disjunctivism*. Oxford: Oxford University Press.
- Reins, Louisa M., and Alex Wiegmann. 2021. 'Is Lying Bound to Commitment? Empirically Investigating Deceptive Presuppositions, Implicatures, and Actions'. *Cognitive Science* 45. <https://doi.org/10.1111/cogs.12936>.
- Sartre, Jean-Paul. 1956. *Being and Nothingness: An Essay on Phenomenological Ontology*. Washington Press.
- Saul, Jennifer. 2000. 'Did Clinton Say Something False?' *Analysis* 60 (3): 255–57.
- Saul, Jennifer. 2012. *Lying, Misleading, and The Role of What Is Said*. Oxford: Oxford University Press.
- Schroeder, Mark. manuscript. 'Knowledge Is Not the Most General Factive Stative Attitude'.
- Siegler, Frederick A. 1966. 'Lying'. *American Philosophical Quarterly* 3 (2): 128–36.
- Sosa, Ernest. 1999. 'How to Defeat Opposition to Moore'. *Philosophical Perspectives* 33 (13): 141–53. <https://doi.org/10.1111/0029-4624.33.s13.7>.
- Stokke, Andreas. 2018. *Lying and Insincerity*. Oxford: Oxford University Press.
- Trpin, Borut, Anna Dobrosovstnova, and J G Sebastian. 2020. 'A Computer Simulation Study of Graded Lies and Trust Dynamics'. *Synthese*. <https://doi.org/10.1007/s11229-020-02746-5>.
- Turri, Angelo, and John Turri. 2015. 'The Truth about Lying'. *Cognition* 138: 161–68. <https://doi.org/10.1016/j.cognition.2015.01.007>.
- Turri, Angelo, and John Turri. 2019. 'Lying, Fast and Slow'. *Synthese*.
- Turri, John. 2010. 'Does Perceiving Entail Knowing?' *Theoria* 76 (3): 197–206. <https://doi.org/10.1111/j.1755-2567.2010.01065.x>.
- Turri, John. 2016. *Knowledge and the Norm of Assertion*.
- Turri, John. 2017. 'Experimental Work on the Norms of Assertion'. *Philosophy Compass* 12 (7): 1–9. <https://doi.org/10.1111/phc3.12425>.
- Weissman, Benjamin, and Marina Terkourafi. 2018. 'Are False Implicatures Lies? An Experimental Investigation'. *Mind & Language*, 34, 221–246. <https://doi.org/10.1111/mila.12212>.
- Whyte, Jamie. 2013. 'Review of "Lying, Misleading & What Is Said"', by J. M. Saul'. *The Philosophical Quarterly*, 209–10.
- Wiegmann, Alex, and Jörg Meibauer. 2019. 'The Folk Concept of Lying'. *Philosophy Compass*, no. April: 1–11. <https://doi.org/10.1111/phc3.12620>.
- Wiegmann, Alex, Ronja Rutschmann, and Pascale Willemsen. 2017. 'Empirically Investigating the Concept of Lying'. *Journal of Indian Council of Philosophical Research* 34 (3): 591–609. <https://doi.org/10.1007/s40961-017-0112-z>.
- Wiegmann, Alex, Jana Samland, and Michael R. Waldmann. 2016. 'Lying despite Telling the Truth'. *Cognition* 150: 37–42. <https://doi.org/10.1016/j.cognition.2016.01.017>.
- Wiegmann, Alex. forthcoming. True lies and Moorean redundancy. *Synthese*.
- Williams, Bernard Arthur Owen. 2002. *Truth and Truthfulness An Essay in Genealogy*. Princeton: Princeton University Press.
- Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford: Oxford University Press. <https://doi.org/10.5040/9781474284974.00620>.