

Counterevidentials

Laura Caponetto¹ and Neri Marsili²

Moorean constructions are famously odd: it is infelicitous to deny that you believe what you claim to be true. But what about claiming that p, only to immediately put into question your evidence in support of p? In this paper, we identify and analyse a class of quasi-Moorean constructions, which we label counterevidentials. Although odd, counterevidentials can be accommodated as felicitous attempts to mitigate one's claim right after making it. We explore how counterevidentials differ from lexicalised mitigation operators, parentheticals, and anaphoric mitigation devices, and consider some cognate non-assertoric constructions. We conclude by exploring the implications of our analysis for theorising about linguistic responsibility, assertion, and lying.

1. Moorean Assertions

Moore (1942: 542–43, 1944: 204) famously observed that it would be ‘perfectly absurd or nonsensical’ for someone to assert a statement like (1) or (2):

- (1) Mulberries grow on trees, but I don’t believe that they do.
- (2) My bicycle is parked under the bridge, but I don’t know that it is.

These utterances *feel contradictory* and *sound paradoxical*. However, as Moore notes, both (1) and (2) could easily be true. The speaker—call her Alice—may believe that mulberries don’t grow on trees, even if in fact they do. And Alice’s bicycle may have been stolen after she parked it, in which case she would believe incorrectly (and thus fail to know) that it is parked under the bridge. Even if (1) and (2) sound paradoxical, they involve no contradiction.

Albeit non-contradictory, Moorean assertions like (1) and (2) lead to a ‘catastrophic’ *illocutionary failure*: presented with (1), a competent speaker of English (no matter how charitable) wouldn’t be able to make sense of what Alice is trying to communicate. We can’t really assume that Alice is trying to claim that mulberries

¹ University of Cambridge, lc882@cam.ac.uk

² UNED, neri@fsf.uned.es

grow on trees, since she goes on to deny that she believes it; and we can't really assume that she is trying to communicate that she does not believe it, since she stated that they do grow on trees. As a result, Alice's assertion misfires.

There are several competing explanations for the perceived paradoxicality of Moorean assertions.³ According to one influential view, we perceive Moorean assertions to be paradoxical because the sincerity conditions of one conjunct are denied in the other, in such a way that the speaker performs a speech act only to immediately admit that it is infelicitous. This generates the sense of oddness that is typical of Moorean assertions,⁴ and results in a self-defeating illocution, that is, an illocution that cannot be successfully performed due to having contradictory success conditions (Vanderveken 1980: 249; Searle and Vanderveken 1985: 91).

Denying sincerity conditions is not the only way to generate Moorean paradoxicality. The same pragmatic phenomenon can be triggered by an overt denial that some *other* felicity conditions for the performance of a given assertion are met. In addition to sincerity, a fully felicitous assertion requires that the speaker have some reasons (grounds, evidence) in support of the proposition they are putting forth.⁵ Denying that one has such reasons gives rise to a Moorean⁶ feeling. Consider the following utterance:

- (3) Mel broke your trophy, but I don't have any reason to believe that Mel broke your trophy.

³ See Williams (2015) for a general overview. See also Green and Williams (2007: chs. 1–2) for a good introduction to Moore's Paradox.

⁴ Some authors note that, in a few exceptional cases, Moorean assertions need not be completely absurd (Crimmins 1992; Hájek and Stoljar 2001; Clark 2002: 118–20; Rosenthal 2002; Gallois 2007: 166; Chan 2008; Pruss 2012; Labinaz 2022). For our purposes, however, these exceptional cases do not matter: what matters is that *typically* Moorean constructions are fatally infelicitous.

⁵ Here we are focusing on felicity, rather than conformity to the putative 'norm of assertion' that, according to Williamson (2000), uniquely regulates this speech act. However, it should be noted that speech act theorists and epistemologists alike recognise that unwarranted assertions are defective, even when they are believed to be true (for an overview, see Pagin and Marsili 2021: §5.1).

⁶ We will use the label 'Moorean' in a liberal way throughout this paper, to cover non-classic constructions as well as classic ones. In doing so, we do not aim to take a stance on whether these constructions are genuinely Moorean, nor are we aiming to deny that (3) is significantly different from (1) and (2). Our sole purpose is to highlight some important similarities between these constructions: their apparent paradoxicality, and their illocutionary infelicity.

(3) sounds paradoxical, although no proper contradiction is involved. Like classic Moorean constructions, (3) is a self-defeating speech act: the right conjunct explicitly *denies* that a felicity condition for the left conjunct (i.e. that the speaker has reasons to believe what they asserted) is satisfied, making the speaker's overall illocution fall flat.

2. Counterevidentials

In (3) the speaker denies having *any reason at all* in support of their initial statement. We are not the first to note that such a denial leads to some sort of Moorean paradoxicality (Rudy Hiller 2016: 42; Woods 2018: 335; Littlejohn 2020: 712). A much less explored question is whether acknowledging that the reasons in support of your assertion are not all that reliable leads to a similar sort of infelicity. Consider the following case, adapted from Arico and Fallis (2013: 802):

- (4) JILL (to Jack): Mel broke your trophy—but I was drunk last night, and there were lots of people in there.

Or the following, from Fallis (2009):

- (5) BEPPO: Tony was with me at the time of the murder. Of course, you know I am really bad with dates and times.

Assertions like (4) and (5) are instances of what we term *counterevidentials*. While counterevidentials have received some attention in the literature on lying (see Fallis 2009; Carson 2010: 38; Arico and Fallis 2013; Fallis 2013; Stokke 2017; Timmermann and Viebahn 2020; Marsili 2021), their striking similarity with Moorean constructions has so far been neglected, and little has been done in the way of attempting to explain their inner workings. This paper aims to fill this gap.

We begin by examining the structure and communicative function of counterevidentials (§2), and then identify their distinctive features vis-à-vis Moorean assertions and other cognate constructions (§3). After exploring whether counterevidential utterances can occur in non-assertoric discourse (§4), in the second part of the paper we move on to draw out the implications of our analysis for philosophical theories of assertion, lying, and linguistic responsibility. We show that counterevidentials represent a novel datapoint for testing theories about the

norm of assertion, complementing existing tests based on Moorean constructions. In particular, they offer new insights into the graded nature of assertability, and how it can be accommodated by different accounts of assertion (§5.1). Counterevidentials have also been invoked to challenge *responsibility-based* definitions of lying. However, we argue that they only pose a threat to views that rely on ‘polar’ accounts of commitment (according to which either one commits to truth, or one doesn’t—no degrees are allowed). This provides further support for understanding *assertoric commitment* as a graded notion—a conclusion that has broader consequences for the pragmatics and epistemology of communication (§5.2).

2.1. What Is a Counterevidential?

Like Moorean assertions, counterevidentials are constituted by two utterances: a plain assertion (‘Mel broke your trophy’, ‘Tony was with me at the time of the murder’), followed by a ‘proviso’⁷ (i.e. a qualification), which conveys that the speaker has poor evidence for what they have just asserted (‘I was drunk last night, and there were lots of people in there?’; ‘you know I am really bad with dates and times?’). Schematically, they have the structure:

$$\Phi \wedge PE_S(\Phi)$$

Here, Φ is the speaker (S)’s assertion, and $PE_S(\Phi)$ the proviso that conveys that S has Poor Evidence (PE) for Φ .⁸

We take it that S has ‘poor evidence’ in the relevant sense whenever S’s evidence falls below the epistemic standard required to assert the first conjunct. Strictly speaking, then, a counterevidential is an assertion followed by a proviso that conveys

⁷ We follow Fallis (2009), who refers to insincere assertions like (4) and (5) as ‘proviso lies’. The term ‘proviso’ refers to the utterance by means of which the speaker qualifies their own grounds for asserting the first conjunct.

⁸ The formula offered here is only meant to help visualise the structure of counterevidential constructions. It is not an attempt to offer a rigorous formalisation of these expressions, which (due to their multifarious nature) would not easily admit such a treatment. Relatedly, the logical operator ‘AND’ is here adopted to refer (quite loosely) to any expression that can play its truth-conditional functions, including adversative conjunctions such as ‘but’.

that the speaker has *poorer evidence than their initial assertion requires*.⁹ To spare the reader convoluted constructions, however, in what follows we will adopt the simpler expression ‘poor evidence’.

While Moorean assertions are virtually non-existent in ordinary language, counterevidentials are attested in everyday exchanges. Here are two examples—the former from the movie *Under Suspicion* (1991) by Simon Moore and the latter from an online forum:

- (6) It was definitely a man, from their clothes, but it was dark. I couldn’t see them very well.
- (7) The St. Patrick’s weekend items are not available either and they should be. Although I haven’t checked in a while.¹⁰

Like the Moorean (3), examples (4)–(7) display an internal tension between their conjuncts. However, while in (3) the speaker admits that they have no evidence *at all* to back up what they said, in (4)–(7) the speakers only *partially* put their evidence into question.¹¹ This points to a substantial difference between Moorean constructions and counterevidentials. As noted in §1, Moorean assertions are fatally

⁹ We are here assuming (as is commonplace in the literature) that epistemically permissible assertions have to meet a certain evidential standard. We are not, however, taking a stance on the (much more controversial) issue of what this evidential standard might be (although we do rule out some candidates in §5.1; for a broader overview, see Pagin and Marsili 2021: §5.1). Importantly, our graded characterisation of ‘poor evidence’ allows for stronger and weaker clashes between the initial assertion and the proviso. Had Beppo uttered:

(5*) Tony was with me at the time of the murder. Of course, I’m occasionally forgetful with dates and times,

his utterance would have still been classified as a counterevidential by our definition—albeit a less Moorean-sounding one than (5). We are grateful to an anonymous reviewer for inviting us to discuss these subtler cases.

¹⁰ ‘Items meant to be available for diamonds’, Glu Communities, January 2022 <<https://communities.glu.com/kim-kardashian-hollywood/discussion/comment/7061984>> (07/10/2022). Most of the counterevidentials that we have identified in the literature and in corpora feature a contrastive conjunction (‘but’, ‘although’). While this isn’t an invariable linguistic feature of counterevidentials (in (5), the proviso is prefaced by the sentence adverb ‘of course’), contrastive conjunctions presumably contribute to making counterevidentials acceptable. Compare ‘Mel broke your trophy—but I was drunk last night’ and ‘Mel broke your trophy—and I was drunk last night’. The former is easier to interpret as an amendment: by conveying a contrast between the left and right conjuncts, the speaker suggests that there is something odd with their speech act, which the hearer is called upon to solve.

¹¹ In (6), the speaker puts into question her ability to see the man; in (7), that the catalogue they checked contained information that is still valid.

infelicitous: one conjunct defeats the other, meaning that there is no charitable way to make sense of what the speaker is trying to communicate. The same is not true of counterevidentials.¹² In the case of counterevidentials, the right conjunct, $PE_s(\Phi)$, only partially puts the felicity of the left conjunct, Φ , into question. The resulting tension is not fatal: there is a robust sense in which the speaker can communicate their claim, and in which the hearer can successfully interpret what they meant. Unlike Moorean assertions, counterevidentials allow for an interpretation that resolves the tension between their conjuncts.

2.2. Counterevidentials as Amendments

A natural way for the hearer to accommodate (4)–(7) is to read the second conjunct as an attempt to adjust or *amend* the first. ‘Amendment’ is a technical term that we borrow from Caponetto (2020): it refers to any communicative attempt to alter the degree of strength of a previously performed speech act—to weaken or strengthen its normative fallout. For example, (8) is an assertion amended into a guess:

(8) Thomas is coming tonight. At least, I guess so.

The qualification ‘At least, I guess so’ demotes the speaker’s initial assertion into a guess—that is, a weaker assertive act that generates a weaker commitment to the truth of the proposition expressed (*that Thomas is coming tonight*), and which requires weaker supporting evidence for appropriate performance (Searle and Vanderveken 1985; Green 2013, 2017; Labinaz and Sbisà 2014).¹³

In amending a previous illocution, a speaker undoes it only *partially*: they rework its normative upshot, rather than wiping it out completely. This makes amendment importantly different from retraction. Retraction *cancel*s a certain act’s normative effects, whereas amendment *alter*s those effects. Once an assertion that p is retracted,

¹² We are not denying that there is a sense in which counterevidentials can be infelicitous. All we are claiming is that they are not *fatally* infelicitous—that is, infelicitous in the same way as Moorean assertions.

¹³ Speakers can also engage in *propositional* amendment: they can (and often do) change what they say, without tweaking the illocutionary force of their performance, as in (8*):

(8*) Thomas is coming tonight. I mean, tomorrow night.

Since this paper focuses on *illocutionary* amendment (precisely, on attempts to downgrade—rather than boost—one’s previous illocution), henceforth, unless otherwise specified, we will use ‘amendment’ in this restricted sense.

the speaker is no longer committed to p (or to provide reasons in favour of p , if challenged). By contrast, to amend an assertion is to alter the degree to which one is committed to its content.

We suggest that counterevidentials, albeit odd, can be accommodated as cooperative contributions to the conversation: the speaker might regard the second conjunct $PE_s(\Phi)$ as an indirect attempt to amend the first conjunct Φ . Take (4). Jill may have realised that she was drunk *right after* claiming that Mel broke the trophy. Upon realising that she was not in a position to claim that Mel broke the trophy, she adds a qualification that clarifies just that—that she was not in a position to make an unqualified assertion. So construed, Jill’s proviso (‘I was drunk last night, etc.’) invites her audience to reinterpret her initial speech act as something weaker: a suggestion, a guess, or perhaps a guarded assertion. If Jack accommodates her manoeuvre, Jill’s assertion gets *amended* into a weaker assertive, and her commitment to the truth of her initial claim (‘Mel broke your trophy’) is reduced. A parallel point can be made about Beppo: we can make sense of (5) by interpreting the second conjunct (‘Of course, you know I am really bad with dates and times’) as an indirect attempt to amend the first (‘Tony was with me at the time of the murder’). So understood, (5) is still a peculiar construction, but need not lead to the fatal illocutionary failure that we see in Moorean constructions such as (1)–(3).¹⁴

Generalising, a counterevidential is a plain assertion (Φ) accompanied by a quasi-Moorean qualification ($PE_s(\Phi)$), which signals that the speaker has poor reasons in support of what they have said.¹⁵ This generates a pragmatic clash between its two

¹⁴ This analysis bears some similarities with McCready (2015: ch. 5)’s discussion of how disclamation works. McCready offers an insightful formal account of how some cases of disclamation (such as ‘I drank the coffee, though it might not really have been coffee’), which would be Moorean if interpreted literally, are accommodated as informative hedges. However, there are two important differences between our analysis and McCready’s. First, McCready is concerned with a very different phenomenon: mitigating expressions that target the matrix clause explicitly and directly. Counterevidentials, by contrast, only *indirectly* put into question their first conjuncts. Second, while the qualifying conjunct of a counterevidential is indisputably asserted, McCready’s solution crucially involves treating the qualifying conjunct of hedges as unasserted, not-at-issue content.

¹⁵ This characterisation makes constructions such as (9) and (10) slightly different from counterevidentials:

(9) Fred wants to go to college. Don’t take my word for it, ask him. (From the 1991 drama *Little Man Tate*.)

conjuncts. But, while genuine Moorean assertions invariably lead to illocutionary failure, counterevidentials can be accommodated as cooperative contributions to a conversation. We have suggested that this is because the tension between their conjuncts can be meaningfully resolved: the second conjunct can be interpreted as a cooperative attempt to downgrade the strength of one's initial assertion. In Caponetto's terminology, this constitutes an *indirect amendment* of the first conjunct. Here 'indirect' indicates that the speaker's intention to downgrade the first conjunct is communicated *implicitly*, by questioning that certain conditions for its felicity are met. A *direct* amendment, by contrast, relies on a lexicalised operator ('I guess', 'perhaps', etc.) that *explicitly* targets the strength of the speaker's speech act.

3. Counterevidentials and Mitigation Devices

On the account developed so far, the communicative function of counterevidential provisos is analogous to that of more familiar *mitigation devices*, such as *evidentials* ('apparently', 'I've heard') or *epistemic modals* ('perhaps', 'maybe'). In assertoric utterances, the function of these expressions is often to reduce the speaker's commitment to the truth of the asserted proposition (Lyons 1977: 797; Holmes 1984; Coates 1987; Caffi 1999; Sbisà 2001, 2014; Benton and van Elswyk 2020; Murray 2021). To illustrate, (12) and (13) are less committal than (11):

(11) Mel broke your trophy.

(12) From what I've heard, Mel broke your trophy.

(13) Perhaps Mel broke your trophy.

There are several obvious differences between (12)–(13) and counterevidentials. A crucial one concerns their dynamics: while in counterevidentials the speaker first

(10) Albert is in his office, but see for yourself; don't take my word for it. (Adapted from Harman 1986: 50.)

The right conjuncts of (9) and (10) do not say anything about the speaker's reasons in support of their assertion. Rather, they suggest that the speaker lacks the perlocutionary intention that is characteristically associated with asserting—namely, the intention to make the hearer believe that *p* partly because they have asserted that *p*. Since one can felicitously assert without having this intention, (9) and (10) sound much more natural, and less odd, than counterevidentials. Meanwhile, the neo-Gricean idea that one *cannot* assert without attempting to convince one's audience is subject to known objections, which many scholars regard as fatal. See, esp., Searle 1969: 44–47; Vlach 1981; Alston 2000: 44–50; Glüer and Pagin 2003; Siebel 2003, 2020; Green 2007: 75–82; MacFarlane 2011: 80–84; García-Carpintero 2018; Marsili 2020b.

makes an assertion and then attempts to (indirectly) amend it, here the speaker mitigates their assertion *as* they make it. In (12), for instance, the speaker first announces that the assertion they are going to make is about what they have heard (as opposed to how things are), and then makes it. No amendment is at work, since one can only amend previously performed speech acts—not speech acts that one is about to make.¹⁶

Mitigation operators can also appear as *sentence-final* parentheticals, to weaken an unmitigated assertion just *after* it is made, as in the following examples:

(14) Mel broke your trophy, I think.

(15) Mel broke your trophy, I've heard.

The speakers of (14) and (15) at first appear to be making full-fledged assertions, but they then proceed to qualify their statements with a parenthetical ('I think', 'I've heard').¹⁷ The addition of the parenthetical leads the speaker to ultimately undertake only a partial, guarded commitment to the truth of the proposition expressed by the matrix sentence ('Mel broke your trophy'). Insofar as they downgrade commitment, sentence-final parentheticals and counterevidential provisos play an analogous communicative function. The analogy with mitigation devices is even more striking if we consider constructions in which mitigation is not achieved via a parenthetical, but via a second 'standalone' assertion that amends the original one. Consider the following utterances:

(16) Mel broke your trophy. Perhaps he did.

(17) Mel broke your trophy. Probably he did.

¹⁶ Also, while the qualifiers 'from what I've heard' and 'perhaps' mitigate whatever content comes under their scope, placing a counterevidential qualification at the beginning of an assertion does not have quite the same effect. Consider:

(4*) I was drunk last night, and there were lots of people in there. But Mel broke your trophy.

As noted by a referee, the most natural interpretation of (4*) is to take Jill to be suggesting that her evidence is so conclusive that she is willing to make a flat-out assertion even if she was in an inebriated state at the time of the events (imagine, for instance, that Jill unequivocally saw Mel smash the trophy onto the floor). Since the qualification at the beginning of (4*) does not have the function of downgrading the speaker's main claim, we do not regard constructions like (4*) as genuine counterevidentials.

¹⁷ There is disagreement about the exact mechanisms through which parenthetical mitigation is achieved (see Urmson 1952; Ross 1973; Asher 2000; Kärkkäinen 2003; Simons 2007; McCready 2015; Murray 2021), but such details are irrelevant to our purposes.

(18) Mel broke your trophy. Allegedly he did.

Call this particular kind of mitigation *anaphoric mitigation*. Like counterevidentials, (16)–(18) display an internal tension: a plain assertion is made, and then repeated,¹⁸ but in a mitigated way. The performance of two assertions with the same content but different strengths in succession generates a sense of oddness, which can be resolved by interpreting the second utterance as an attempt to amend the first—that is, by taking the speaker to be clarifying that they only mean to commit to the second, mitigated claim.

The analogy between anaphoric mitigation and counterevidentials helps to elucidate the pragmatic import of the latter. Much like anaphoric mitigation, counterevidentials are utterance pairs that exhibit an apparent internal clash which can be solved by taking the speaker to be amending their former assertion by means of the latter. The analogy also shows that the accommodation mechanisms invoked in the interpretation of counterevidentials are fairly familiar ones, analogous to those involved in the interpretation of anaphoric mitigation. The main difference between them lies in how the amendment of the first illocution is achieved. In (16)–(18) the amendment is *direct*: it is achieved by a lexicalised operator that clarifies what the amended assertion is (e.g. ‘Perhaps, *p*’, ‘Allegedly, *p*’). In counterevidentials, the amendment of the initial assertion is instead achieved *indirectly*: the amended message thus needs to be worked out by the audience, and is open to interpretation.¹⁹

Now that we have a clearer idea of how counterevidentials differ from cognate constructions, and of the inferential mechanisms through which they are processed, let us move on to consider whether non-assertoric constructions, too, can give rise to counterevidentials.

4. Counterevidential Provisos and Non-assertoric Speech Acts

It is well known that non-assertoric constructions can display Moorean paradoxicality (see Black 1952: 32–33; Searle and Vanderveken 1985; Harnish 2007;

¹⁸ We are using ‘repeated’ liberally, assuming that pronouns are saturated and ellipses filled.

¹⁹ In this, counterevidentials are analogous to *figleaves* (Saul 2017, forthcoming): non-lexicalised expressions that are used to partially undermine a preceding speech act that is racist, sexist, or otherwise problematic.

Kaufmann 2016; Harris 2017; Woods 2018; Mandelkern 2019). To illustrate, imagine that Alice utters one of the following sentences:

- (19) I swear that I locked the car, although I believe I didn't.
- (20) I won't drink at the party, I promise—but I intend to do it.
- (21) Condolences for your loss, although I rejoice at your grief.
- (22) I advise you to climb that wall, but I don't think it's a good idea for you to do it.

Just like Moorean assertions, these utterances *sound paradoxical*, *feel contradictory*, and lead to *illocutionary failure*. However, each pair of conjuncts can simultaneously be true. It could be true that Alice locked the car (and that she is swearing that she locked the car),²⁰ but she may have forgotten that she did, and believe that she didn't. Similarly, it could be true that Alice intends to drink at the party, but also that she will end up not drinking, just as she promised (because, say, there are no drinks at the party). Analogous points can be made about (21) and (22).

Moorean paradoxicality, then, stretches beyond clear-cut assertoric constructions. Does the same hold for counterevidentials? It would seem so. Consider, for instance, (23) and (24).

- (23) I warn you that Mel broke your trophy—but I was drunk last night, and there were lots of people in there.
- (24) I should inform you that Mel broke your trophy—but I was drunk last night, and there were lots of people in there.

These constructions behave just like standard counterevidentials, even though the left conjunct features a performative verb other than 'assert'.²¹ In a way, this is unsurprising, since 'warning' (as used here) and 'informing' are assertive speech acts.

²⁰ There is disagreement about the semantics of explicit performatives. Given an explicit performative of the form 'I hereby *f*', descriptivist theorists hold that its content is 'I hereby *f*' (Hedenius 1963; Lewis 1970; Bach 1975; Bach and Harnish 1979, 1992; Ginet 1979), whereas non-descriptivist theorists hold it to be simply *f* (Harris 1978; Searle 1989; Reimer 1995; Jary 2007; Marsili 2020a). We adopt a non-descriptivist interpretation of these utterances, but this disagreement is tangential to our discussion. No matter which semantics one favours, no contradiction arises in the examples under consideration.

²¹ If anything, (23) and (24) sound slightly more odd than standard counterevidentials, possibly due to the presence of the explicit performative, which emphasises that the speaker is performing a specific illocutionary act, making the subsequent correction all the more puzzling.

But what happens when a proviso is added to an illocution that belongs to a different taxonomic family, such as *commissives*?

The short answer is that matters get more complicated. We saw that counterevidentials do not lead to a catastrophic illocutionary failure: they can be accommodated as attempts to mitigate the speaker's original illocution. Not every kind of speech act admits illocutionary mitigation, though: for instance, a promise accompanied by a warning that the speaker may not fulfil it is not a promise at all (Scanlon 1990: 208–9; Bratman 1999: 136–37; Fallis 2013: 347; Marsili 2021: 3257). At most, a promise can be downgraded into an altogether different illocution, such as an assertion that expresses a mere intention (cf. Owens 2012: 56). Alternatively, it can be amended into a promise to do something less demanding, that is, into a promise with a different content (cf. Marsili 2021: n. 21).²² But it cannot be mitigated into a *weaker* promise with the same content, since a weak promise ('I promise that *maybe* I will φ '; 'I promise that I will φ , *although I might end up not φ -ing*') is simply not a promise. This tends to be true of commissives more broadly: their illocutionary point is to put the speaker under a certain obligation; any hesitation expressed by means of mitigation devices would undermine this goal.²³ If this is right, then it should be impossible to accommodate 'counterevidential promises' as mitigated promises. And this is exactly what we see in (25):

- (25) I promise that I will wake up at 7am tomorrow, but you know that I am really unreliable in the morning.

The speaker of (25) guarantees that they will wake up at 7am, only to deny that it is likely that they will. This results in a self-defeating promise, rather than a merely

²² The propositional content of promises *can* be mitigated. 'I promise that *I will go to the gym*' can be mitigated into 'I promise that *if you buy me a krapfen, I will go to the gym*' or 'I promise that *I will go to the gym a little*' (cf. Marušić 2017). This kind of mitigation is known as propositional mitigation (or 'bush', in Caffi's terminology: see Caffi 1999) and is to be distinguished from illocutionary mitigation (and amendment). See also fn. 11 above.

²³ Strengthening commissives, by contrast, is generally viable. A promise, for instance, can be strengthened into a *solemn* promise. This is because assuring the hearer that one will φ or expressing confidence in one's ability to φ is compatible with the goal of putting oneself under an obligation to φ . We thank an anonymous reviewer for suggesting that we clarify this point.

mitigated one. Unlike (23) or (24), (25) has a truly Moorean flavour, and is much harder to accommodate than a counterevidential.²⁴

What about *directives* and *declarations*? Can we construct genuine counterevidentials with these illocutions? Consider the following:

(26) I advise you to buy the new X1 notebook—although I've got to admit I'm not much of a computer expert.

(27) You are officially appointed as a ceremony valet. Of course, you will have to check with the maître de chambre first, for we may have too many valets already.

(26) is analogous to standard counterevidentials in structure and function: the speaker first performs a speech act, and then questions whether a felicity condition is met. To felicitously advise, the speaker must have the requisite expertise over the relevant subject matter (Sbisà 2018: 36). The right conjunct of (26) questions the speaker's computer expertise, thus clashing with the advice given in the left conjunct.²⁵ While this clash is somewhat puzzling, it can be accommodated by interpreting the speaker's overall performance as a weakened act of advising. By contrast, (27) cannot be accommodated in this way. Appointments (and declarations more broadly) are 'on/off' speech acts: either one appoints someone or not—'partial' appointments are not a possibility. If (27) can be accommodated at all, it is as a *conditional* appointment: 'If there are vacant valet positions, then you are appointed as a ceremony valet' (cf. Searle and Vanderveken 1985: 157–60). So, although (27) has the same surface structure as counterevidentials, it does not count as one, because the qualification here plays a different function: what is amended, if anything, is the *content* of the illocution, not its *force*.

Determining how much these examples generalise to their respective illocutionary classes (directives, declarations) would require a long detour, and would go beyond

²⁴ That noted, a sufficiently creative hearer could still give a charitable interpretation and read (25) as an attempt to amend the initial promise into a mere expression of intention.

²⁵ Advising also has the preparatory condition that the course of action represented in the proposition is in the addressee's interest (Searle and Vanderveken 1985: 203). If I advise you to buy a certain notebook, but I know little of computers, what I am advising you to do may turn out to run *against* your interests. This further explains the perceived clash between the two conjuncts of (26), and the overall sense of oddness that the utterance generates.

the ambitions of this paper. We shall rest content in noting that genuine counterevidentials only arise with illocutions that give rise to mitigable commitments, for it is only in those cases that the qualifying proviso can be interpreted as an attempt to downgrade the strength of the initial speech act.

5. Implications for Theorising about Assertion, Lying, and Linguistic Responsibility

Moorean assertions have attracted scholarly attention because they have important implications for more general theories—typically, theories of how communication works, and of the normative expectations governing conversation. For instance, Moorean assertions involving denial of belief have been used to support the thesis that the essential effect of making an assertion is to express a belief in its content (Moore 1912: 125; Williams 1982). Similarly, it has been suggested that Moorean assertions involving denial of knowledge sound paradoxical because assertion is governed by a norm dictating that one should not assert what one does not know (Williamson 2000: 253; Shaffer 2012; Littlejohn 2020). Reflecting on counterevidentials can also offer insights into our understanding of how communication works at large. In what follows, we outline some interesting implications for theorising about assertion, lying, and linguistic responsibility.

5.1. Attitude Expression and Norms of Assertion

We have characterised counterevidentials as assertions of the form $\Phi \wedge \text{PE}_s(\Phi)$, displaying a quasi-Moorean clash that can be solved by interpreting the proviso $\text{PE}_s(\Phi)$ as an attempt to downgrade the strength of the initial assertion Φ . A good theory of assertion should be able to explain why the addition of the proviso $\text{PE}_s(\Phi)$ is perceived as odd, and why it is naturally accommodated as an amendment. As we are about to see, not every theory on the market is compatible with this prediction.

Among the theories that *are* able to accommodate this fact is the influential account that takes assertions to express knowledge (see Unger 1975: 256–70; Slote 1979: 179), and/or to be governed by a knowledge norm (i.e. ‘You should assert a proposition only if you know that proposition’: see, e.g., Williamson 2000). On this

view, if you assert that Mel broke a certain trophy, you are communicating that you know that he did, and you should not have made your assertion unless you in fact knew that he did. The envisaged link between knowledge and assertion can accommodate our data because knowledge entails justification: you know a proposition only if you are justified in believing it. In other words, if you have poor evidence in favour of a proposition, you cannot know it and should not assert it. This hypothesis has the advantage of being able to explain what happens in counterevidentials such as (4): when Jill admits that she has poor evidence in support of her claim (she was drunk, and way too many people were present), she informs the hearer that she is not in a position to know that what she was saying is true. The resulting tension with the knowledge norm explains why the utterance is somewhat odd, and why accommodating (4) as a cooperative contribution requires interpreting it as an attempt to downgrade the illocution into a weaker assertive. Since assertions must be known, Jill's proviso can be regarded as cooperative only when interpreted as an attempt to clarify that, not being in a position to *know*, she wants her initial claim to be read as something weaker than assertion: a guess, or a suggestion.²⁶

Now consider two rival views. First, the view that assertion expresses belief and is governed by a belief rule (i.e. 'Assert a proposition only if you believe that it is true': see, e.g., Bach 2008; cf. also Hindriks 2007). Second, the view that assertion is governed by a truth rule (i.e. 'Assert a proposition only if that proposition is true': see, e.g., Weiner 2005; Whiting 2012). These views establish no direct link between assertion and justification (or evidence). You can make a perfectly appropriate assertion even if you do not have any evidence in its support, as long as you believe it (according to the former view), or as long as it is true (according to the latter).

If we assume either of these two views, counterevidentials become somewhat more mysterious. It is not clear why we should feel a tension between their two conjuncts, and it is unclear why interpreting (4) as cooperative would require us to take it to be an amended assertion. To illustrate, let's focus on the first view: if assertions express beliefs and are governed by a belief rule, in uttering (4), Jill should be simply and

²⁶ A similar explanation can be generated from the hypothesis that assertions are governed by a justification norm that requires that speakers only assert what they are justified to believe (or something of the sort). This view has been defended by various authors, including Douven 2006; Lackey 2007; Kvanvig 2009; Gerken 2012, 2017; Kneer 2018; Reuter and Brössel 2019.

unproblematically communicating that she believes that Mel broke the trophy, and that she holds this belief despite the fact that the evidence supporting her belief is poor. This is in principle a cooperative contribution on the belief view. This account has no obvious resources with which to predict a tension between the first claim and the counterevidential qualification, nor to explain why it is natural to regard Jill's proviso as an attempt to mitigate her initial assertion. A parallel point can be made about the truth rule.

To be sure, we are not suggesting that proponents of the belief rule (or the truth rule) cannot derive from their account a slightly more complex story that accommodates the data.²⁷ What we are suggesting is rather that, much like Moorean assertions, counterevidentials offer a fertile terrain for discussion. There has been much controversy concerning which account of assertion is in the best position to explain how Moorean constructions (in all their variety) behave.²⁸ Here we are noting that counterevidentials invite a similar discussion, and that they may represent a useful datapoint against which to test the explanatory power of competing theories about the norm of assertion, as well as illocutionary normativity more generally.

Importantly, counterevidentials represent a *novel* datapoint, not reducible to comparable Moorean constructions already discussed in the literature, such as (3) ('Mel broke your trophy, but I don't have any reason to believe that Mel broke your trophy'). This is for various reasons. First, Moorean assertions like (3) involve a binary contrast between conjuncts (the right conjunct denies that the speaker has *any* supporting reasons or evidence); counterevidentials, however, concern the *quality* of the evidence available to the speaker, a factor that comes in degrees (the right conjunct conveys *how good* the speaker's evidence is). Those who conceive of assertability as a graded notion (Gerken 2012, 2014, 2017; Carter 2022) might thus find counterevidentials more interesting than Moorean assertions to test the

²⁷ In this respect, it may help to appeal to a distinction between primary and secondary propriety, as developed by DeRose 2002; see also Weiner 2005; Whiting 2012. Note, however, that this solution is subject to controversy, and several scholars regard the distinction as spurious (Douven 2006: 478–80; Lackey 2007; Gerken 2011; Cappelen 2011: 46; Pagin 2016; Schechter 2017; Marsili 2018: 645–46).

²⁸ See Williamson 2000: 253; Weiner 2005: 253–54; Douven 2006: 474–75, 2009; Lackey 2007; Chan 2008; Littlejohn 2010, 2020; Cappelen 2011: 38–40; Shaffer 2012; Woods 2018.

explanatory powers of their theories. Second, counterevidentials concern a different linguistic phenomenon than Moorean assertions, testing different explanatory virtues of a given theory of assertion. While a theory of Moorean assertions must explain the catastrophic oddity generated by their conjuncts, an account of counterevidentials must instead explain their internal dynamic—that is, how one conjunct may successfully amend the other. Lastly, unlike Moorean assertions, counterevidentials can be meaningful contributions to a conversation, and naturally occur in ordinary language. This difference underscores their significance as a more ‘ecologically valid’ test for norms of assertion, reflecting real-world language use.

Incorporating counterevidentials into the set of tests for norms of assertion, in sum, allows for a more thorough and ecologically valid assessment of competing theories, improving our understanding of the principles governing the making and amending of assertoric speech acts.

5.2. Lying and Assertoric Commitment

As mentioned at the outset, constructions akin to counterevidentials have already received some attention in the philosophical literature, within debates on lying (see Fallis 2009; Carson 2010: 38; Arico and Fallis 2013; Fallis 2013; Stokke 2017; Timmermann and Viebahn 2020; Marsili 2021) and assertoric commitment (see Rudy Hiller 2016: 49–51; Marsili 2020b: §5, Marsili forthcoming).

In the debate on lying, counterevidentials are taken to pose a challenge for *responsibility-based views*, which hold that lying requires undertaking responsibility for the truth of what one has communicated (see Carson 2006, 2010; Saul 2012; Marsili 2014, 2021; Leland 2015; Viebahn 2021). Intuitively, counterevidentials can be lies (and are known in the literature as ‘proviso lies’): if Jill knows that she is the one who broke the trophy and utters (4) (‘Mel broke your trophy—but I was drunk last night, and there were lots of people in there’), Jill is intuitively lying.²⁹ But (the argument goes) the proviso in (4) deprives Jill’s initial utterance (‘Mel broke your trophy’) of the commitment it would have generated otherwise. Therefore, Jill is

²⁹ This intuition is shared by most laypeople, as shown by Arico and Fallis (2013).

lying without taking responsibility for the truth of what she says—*responsibility-based* accounts of lying must be false (Fallis 2009, 2013; Arico and Fallis 2013).

This argument assumes that counterevidential qualifications void or cancel the responsibilities that the speaker undertook with their initial assertion. Our analysis of counterevidentials says otherwise. Provisos do not *cancel* the assertion they target: they merely *amend* it, reducing the strength of the speaker's commitment to their initial claim. So understood, 'proviso lies' present no challenge to *responsibility-based* accounts of lying: Jill is still committed to the truth of the first conjunct of (4), albeit in a mitigated way.

That noted, the objection retains its plausibility when applied to those responsibility-based accounts of lying that deny that discursive responsibility comes in degrees (Carson 2006, 2010; Viebahn 2021). According to these views (and to some accounts of discursive commitment—cf. Katriel and Dascal 1989; Geurts 2019), *undertaking assertoric responsibility* is an on/off predicate: either the speaker undertakes responsibility, or not—no degrees are allowed.³⁰ This rules out the possibility of amending an assertion, because amendment presupposes that one can decrease one's commitment to *p* without cancelling it. For these authors, provisos cannot *downgrade* the speaker's original assertion: they either cancel it, or have no effect at all.³¹

This is clearly an unpalatable conclusion. *Pace* these 'polar' accounts, assertion does not behave like speech acts that give rise to on/off responsibilities. Declarations (and to some extent commissives: cf. §4) generate responsibilities of the on/off kind: their force cannot be downgraded by adding a counterevidential qualification.

³⁰ These authors use different terms to refer to the relevant responsibilities: Carson talks about *warrant*, Viebahn about *commitment*. Relatedly, they have different reasons to regard 'undertaking responsibility' as a polar predicate. For Carson (2010), this is because warrant is tied to promissory commitment (Carson disputes this, but see Fallis 2013 and Marsili 2021 for discussion of why Carson is committed to this thesis); for Viebahn (2021), it is because commitment is tied to the notion of consistency, which does not allow for degrees (Viebahn 2021: n. 46).

³¹ One might reply that polar accounts ultimately allow for degrees, understood as graded differences in what the speaker is committed *to doing*. For instance, one could claim that, while an unmitigated assertion fully commits you to *knowing* that *p*, a mitigated assertion fully commits you to *reasonably believing* that *p*. This solution would indeed be available, but authors such as Viebahn (2021: 304) explicitly deny it (any commitment falling short of a commitment to *knowing* the proposition is not assertoric in his view). The upshot is that genuine polar accounts (such as Viebahn's) take commitment not to allow for degrees, even in this weak sense.

If these authors were right, counterevidential assertions would generate the same sort of catastrophic infelicity. Clearly this is not the case: unlike proviso promises (such as (25)) and proviso declarations (such as (27)), Jill's (4) can easily be accommodated as a mitigated assertion. Polar accounts misrepresent the kind of responsibilities that a speaker undertakes when they make an assertion.

In reply, an alternative analysis of proviso lies could be proposed, which splits the speaker's contribution in two. At time t_1 , the speaker makes a genuine assertion by uttering the first conjunct ('Mel broke your trophy'). Then, at t_2 , the speaker takes back *all* assertoric responsibility (not just part of it) by adding the proviso ('but I was drunk last night, etc.'). This analysis is compatible with a polar understanding of commitment, if complemented with an error theory about why ordinary speakers classify proviso lies as genuine lies. It might be suggested that laypeople's intuitions are thrown off course because the speaker first makes an assertion (which is compatible with lying) and then takes it back (which is incompatible with lying). The internal incoherence displayed by proviso lies may thus blur laypeople's intuitions, leading them to judge—erroneously—that proviso lies are lies. According to this error-theoretic argument, polar accounts make the correct prediction after all: contrary to laypeople's unreliable intuitions, proviso lies are *not* lies, because the speaker is not committed (to any extent) to the first (insincere) conjunct.³²

Albeit ingenious, this reply is open to a known objection (Fallis 2013: 348). The proposed error theory classifies (4) as a mendacious assertion that is immediately and fully taken back. But lies that are retracted (or annulled: cf. Caponetto 2020) immediately after being uttered, such as (28) and (29), are almost invariably regarded as cooperative, non-mendacious contributions to the conversation (cf. Goffman 1974: 87; Fallis 2013: 348; Dynel 2018: §5.3.1):

(28) Mel broke your trophy. Just kidding!

(29) Mel broke your trophy. Actually, let me take that back, it might have been someone else.

³² We thank an anonymous reviewer for suggesting this alternative explanation. For a like-minded suggestion, see Carson (2010: 37–39), to which Fallis (2013: 348) responds.

So, the error theory just proposed would not be enough. If proviso liars fully revoke their commitment to the first conjunct (as the error theorist argues), we would need an explanation of why these cases are different from constructions like (28) and (29). We would need a principled criterion to explain why lies that are immediately taken back (such as (28) and (29)) do not count as genuine lies, *except* when the retraction employs a counterevidential proviso. It's hard to see how such a story could be cashed out without invoking what strikes us as the most obvious difference between the two cases—namely, that proviso lies (as opposed to immediately retracted lies) only retract part of the commitment.

Furthermore, regardless of how plausible the error-theoretic response is, these accounts still face the problem of relying on a polar account of commitment that is problematic for independent reasons. Polar accounts of commitment are unable to make sense of the fact (discussed at length in this article) that counterevidentials can be interpreted as cooperative amendments that downgrade (as opposed to cancel) the commitments undertaken by the first conjunct. By postulating that commitment does not come in degrees, then, these definitions of lying take on board a theory that faces independent difficulties—among them their inability to explain how counterevidentials behave.

Summing up, understanding assertoric force (and insincere assertion, i.e. lying) requires an account of assertoric commitment that allows for varying degrees of strength³³—for only such an account will be able to accommodate the idea that counterevidentials *amend* (rather than cancel) the speaker's responsibility for the truth of the proposition, and the idea that proviso lies involve a *weakened* commitment to the truth of the proposition.

6. Conclusion

Counterevidentials, like Moorean assertions, are a fascinating linguistic phenomenon. While the paradoxicality of Moorean assertions has received a great deal of attention since Moore's seminal work, counterevidentials (and the sense of

³³ Various accounts of illocutionary commitment compatible with gradability have been developed in the literature (see Bazzanella, Caffi, and Sbisà 1991; Sbisà 2001; Labinaz and Sbisà 2014; Krifka 2019; Marsili 2020b). Furthermore, some authors have suggested that the degree of explicitness of an assertion also affects the extent to which the speaker is committed to it (Moeschler 2013; Michaelson 2016: 479–80; Mazzarella *et al.* 2018; Borg 2019; Hall and Mazzarella 2023).

oddness they generate) have been undeservedly ignored by most philosophers and linguists so far. This paper has suggested that this lacuna should be filled. We offered a characterisation of counterevidentials, identifying some important analogies and disanalogies between these constructions and their Moorean cousins. We explored their connections with a variety of cognate constructions, and drew some significant implications for philosophical research on illocutionary normativity, the norm of assertion, the nature of commitment, and the definition of lying³⁴.

References

- Alston, William (2000) *Illocutionary Acts and Sentence Meaning*. Cornell University Press.
- Arico, Adam J., and Don Fallis (2013) ‘Lies, Damned Lies, and Statistics’, *Philosophical Psychology* 26(6): 790–816. <https://doi.org/10.1080/09515089.2012.725977>.
- Asher, Nicholas (2000) ‘Truth Conditional Discourse Semantics for Parentheticals’, *Journal of Semantics* 17(1): 31–50. <https://doi.org/10.1093/jos/17.1.31>.
- Bach, Kent (1975) ‘Performatives Are Statements Too’, *Philosophical Studies* 28(4): 229–36.
- Bach, Kent (2008) ‘Applying Pragmatics to Epistemology’, *Philosophical Issues*, 18: 68–88.
- Bach, Kent, and Robert Harnish (1979) *Linguistic Communication and Speech Acts*. MIT Press.
- Bach, Kent, and Robert Harnish (1992) ‘How Performatives Really Work: A Reply to Searle’, *Linguistics and Philosophy* 15: 93–110. <https://doi.org/10.1007/S10988-01>.
- Bazzanella, Carla, Claudia Caffi, and Marina Sbisà (1991) ‘Scalar Dimensions of Illocutionary Force’, in Igor Žagar, ed., *Speech Acts: Fiction or Reality*: 63–76. Ljubljana: IPrA.
- Benton, Matthew A., and Peter van Elswyk (2020) ‘Hedged Assertion’, in Sanford Goldberg, ed., *The Oxford Handbook of Assertion*: 243–63. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190675233.013.11>.
- Black, Max (1952) ‘Saying and Disbelieving’, *Analysis* 1(2): 25–33.
- Borg, Emma (2019) ‘Explanatory Roles for Minimal Content’, *Nous* 53(3): 513–39. <https://doi.org/10.1111/nous.12217>.

³⁴ This paper was conceived as a joint work between Neri Marsili and Ivan Milić, but after a promising start it was abandoned for several years. During the Covid-19 pandemic, Marsili discussed the core idea of the paper with Laura Caponetto. This work is the result of their subsequent discussions. Although Milić could not re-embark on the project, he made a significant contribution to the paper, especially to its early versions. He also gave us constructive feedback at various stages of the development of the manuscript, for which we are deeply grateful. We would also like to thank Claudia Bianchi and Manuel García Carpintero for their help.

- Bratman, Micheal E. (1999) *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge University Press.
- Caffi, Claudia (1999) ‘On Mitigation’, *Journal of Pragmatics* 31: 881–909.
- Caponetto, Laura (2020) ‘Undoing Things with Words’, *Synthese* 197(6): 2399–414. <https://doi.org/10.1007/s11229-018-1805-9>.
- Cappelen, Herman (2011) ‘Against Assertion’, in Jessica Brown and Herman Cappelen, eds., *Assertion: New Philosophical Essays*: 21–48. Oxford University Press. <https://doi.org/10.1093/acprof>.
- Carson, Thomas (2006) ‘The Definition of Lying’, *Noûs* 40(2): 284–306.
- Carson, Thomas (2010) *Lying and Deception*. Oxford University Press.
- Carter, Sam (2022) ‘Degrees of Assertability’, *Philosophy and Phenomenological Research*, 104(1): 19–49. <https://doi.org/10.1111/phpr.12725>.
- Chan, Timothy (2008) ‘Belief, Assertion and Moore’s Paradox’, *Philosophical Studies*, 139: 395–414. <https://doi.org/10.1007/s11098-007-9130-z>.
- Clark, Michael (2002) *Paradoxes from A to Z*. Routledge.
- Coates, Jennifer (1987) ‘Epistemic Modality and Spoken Discourse’, *Transactions of the Philological Society*, 85(1): 110–31. <http://onlinelibrary.wiley.com/doi/10.1111/j.1467-968X.1987.tb00714.x/abstract>.
- Crimmins, Mark (1992) ‘I Falsely Believe That P’, *Analysis* 52(3): 191.
- DeRose, Keith (2002) ‘Assertion, Knowledge, and Context’, *The Philosophical Review* 111(2): 167–203.
- Douven, Igor (2006) ‘Assertion, Knowledge, and Rational Credibility’, *The Philosophical Review* 115(4): 449–85. <https://doi.org/10.1215/00318108-2006-010>.
- Douven, Igor (2009) ‘Assertion, Moore, and Bayes’, *Philosophical Studies* 144(3): 361–75. <https://doi.org/10.1007/s11098-008-9214-4>.
- Dynel, Marta (2018) *Irony, Deception and Humour*. De Gruyter.
- Fallis, Don (2009) ‘What Is Lying?’, *Journal of Philosophy* 106(1): 29–56.
- Fallis, Don (2013) ‘Davidson Was Almost Right About Lying’, *Australasian Journal of Philosophy* 91(2): 337–53. <https://doi.org/10.1080/00048402.2012.688980>.
- Gallois, André (2007) ‘Consciousness, Reasons, and Moore’s Paradox’, in Mitchell Green and John Williams, eds., *Moore’s Paradox*: 165–88. Oxford University Press.
- García-Carpintero, Manuel (2018) ‘On the Nature of Presupposition’, *Erkenntnis*, 85: 269–293. <https://doi.org/10.1007/s10670-018-0027-3>.
- Gerken, Mikkel (2011) ‘Warrant and Action’, *Synthese* 178(3): 529–47. <https://doi.org/10.1007/s11229-009-9655-0>.
- Gerken, Mikkel (2012) ‘Discursive Justification and Skepticism’, *Synthese* 189(2): 73–94. <https://doi.org/10.1007/s11229-012-0076-0>.
- Gerken, Mikkel (2014) ‘Same, Same but Different’, *Philosophical Studies* 168(3): 725–44. <https://doi.org/10.1007/s11098-013-0156-0>.
- Gerken, Mikkel (2017) *On Folk Epistemology*. Oxford University Press.
- Geurts, Bart (2019) ‘Communication as Commitment Making’, *Theoretical Linguistics* 45(1–2): 1–30.

- Ginet, Carl (1979) 'Performativity', *Linguistics and Philosophy* 3: 245–65.
- Glüer, Kathrin, and Peter Pagin (2003) 'Meaning Theory and Autistic Speakers', *Mind and Language* 18(1): 23–51. <https://doi.org/10.1111/1468-0017.00213>.
- Goffman, Erving (1974) *Frame Analysis*. Northeastern University Press.
- Green, Mitchell (2007) *Self-Expression*. Oxford University Press.
- Green, Mitchell (2013) 'Assertions', in Marina Sbisa and Ken Turner, eds., *Pragmatics of Speech Actions*: 387–410. De Gruyter. <https://doi.org/10.1515/9783110214383.387>.
- Green, Mitchell (2017) 'Assertion', in *Oxford Handbook Online*. <https://doi.org/10.1093/oxfordhb/9780199935314.013.8>.
- Green, Mitchell, and John Williams (2007) *Moore's Paradox*. Oxford University Press.
- Hájek, Alan, and Daniel Stoljar (2001) 'Crimmins, Gonzales and Moore', *Analysis* 61(3): 208–13.
- Hall, Alison, and Diana Mазzarella (2023) 'Pragmatic Inference, Levels of Meaning and Speaker Accountability', *Journal of Pragmatics* 205: 92–110. <https://doi.org/10.1016/j.pragma.2022.12.007>.
- Harman, Gilbert (1986) *Change in View*. MIT Press.
- Harnish, Robert (2007) 'Performative Utterances: Seven Puzzles', *Lodz Papers in Pragmatics* 3(1): 3–21.
- Harris, Daniel (2017) 'Imperatives and Intention-Based Semantics', <https://semanticsarchive.net/Archive/mFiNjhmN/DanielWHarris-ImperativesAndIntentionBasedSemantics.pdf>
- Harris, Roy (1978) 'The Descriptive Interpretation of Performative Utterances', *Journal of Linguistics* 14(2): 309–10. <https://doi.org/10.1017/S0022226700005922>.
- Hedenius, Ingemar (1963) 'Performatives', *Theoria* 29: 115–36.
- Hindriks, Frank (2007) 'The Status of the Knowledge Account of Assertion', *Linguistics and Philosophy* 30(3): 393–406. <https://doi.org/10.1007/s10988-007-9019-5>.
- Holmes, Janet (1984) 'Modifying Illocutionary Force', *Journal of Pragmatics* 8: 345–65.
- Jary, Mark (2007) 'Are Explicit Performatives Assertions?', *Linguistics and Philosophy* 30(2): 207–34. <https://doi.org/10.1007/s10988-007-9015-9>.
- Kärkkäinen, Elise (2003) *Epistemic Stance in English Conversation*. John Benjamins.
- Katriel, Tamar, and Marcelo Dascal (1989) 'Speaker's Commitment and Involvement in Discourse', in Yishai Tobin, ed., *From Sign to Text*: 275–95. John Benjamins.
- Kaufmann, Magdalena (2016) 'Fine-Tuning Natural Language Imperatives', *Journal of Logic and Computation* 29(3): 321–48. <https://doi.org/10.1093/logcom/exw009>.
- Kneer, Markus (2018) 'The Norm of Assertion: Empirical Data', *Cognition* 177: 165–71. <https://doi.org/10.1093/logcom/exw009>.

- Kvanvig, Jonathan (2009) 'Assertion, Knowledge, and Lotteries', in Duncan Pritchard and Patrick Greenough, eds., *Williamson on Knowledge*: 140–60. Oxford University Press.
- Labinaz, Paolo (2022) 'Moorean Utterances and the Illocutionary Dynamics of Assertion', *Intercultural Pragmatics* 19(3): 407–26. <https://doi.org/10.1515/ip-2022-3006>.
- Lackey, Jennifer (2007) 'Norms of Assertion', *Noûs* 41(4): 594–626. <https://doi.org/10.1111/j.1747-9991.2007.00065.x>.
- Leland, Patrick R. (2015) 'Rational Responsibility and the Assertoric Character of Bald-Faced Lies', *Analysis* 75(4): 550–54.
- Lewis, David (1970) 'General Semantics', *Synthese* 22(1–2): 18–67. <https://doi.org/10.1007/BF00413598>.
- Littlejohn, Clayton (2010) 'Moore's Paradox and Epistemic Norms', *Australasian Journal of Philosophy* 88(1): 79–100. <https://doi.org/10.1080/00048400902941240>.
- Littlejohn, Clayton (2020) 'Moore's Paradox and Assertion', in Sanford Goldberg, ed., *The Oxford Handbook of Assertion*: 707–23. Oxford University Press.
- Lyons, John (1977) *Semantics: Volume 2*. Cambridge University Press.
- MacFarlane, John (2011) 'What Is Assertion?', in Jessica Brown and Herman Cappelen, eds., *Assertion: New Philosophical Essays*: 79–96. Oxford University Press.
- Mandelkern, Matthew (2019) 'Practical Moore Sentences', *Noûs* 55(1): 1–23. <https://doi.org/10.1111/nous.12287>.
- Marsili, Neri (2014) 'Lying as a Scalar Phenomenon', in Sibilla Cantarini, Werner Abraham, and Elisabeth Leiss, eds., *Certainty-Uncertainty – and the Attitudinal Space in Between*: 153–73. John Benjamins.
- Marsili, Neri (2018) 'Truth and Assertion: Rules versus Aims', *Analysis* 78(4): 638–48. <https://doi.org/10.1093/analys/any008>.
- Marsili, Neri (2020a) 'Lies, Common Ground and Performative Utterances', *Erkenntnis* 88: 567–78. <https://doi.org/10.1007/s10670-020-00368-4>.
- Marsili, Neri (2020b) 'The Definition of Assertion', <https://doi.org/10.2139/ssrn.3711804>.
- Marsili, Neri (2021) 'Lying, Speech Acts, and Commitment', *Synthese* 199: 3245–69. <https://doi.org/10.1007/s11229-020-02933-4>.
- Marsili, Neri (forthcoming). 'The Definition of Assertion: Commitment and Truth'. *Mind and Language*.
- Marušić, Berislav (2017) 'What's Wrong with Promising to Try?', *Pacific Philosophical Quarterly* 98(S1): 249–56. <https://doi.org/10.1111/papq.12166>.
- Mazzarella, Diana, Robert Reinecke, Ira Noveck, and Hugo Mercier (2018) 'Saying, Presupposing and Implicating', *Journal of Pragmatics* 133: 15–27. <https://doi.org/10.1016/j.pragma.2018.05.009>.
- McCready, Elin (2015) *Reliability in Pragmatics*. Oxford University Press.
- Michaelson, Eliot (2016) 'The Lying Test', *Mind & Language* 31(4): 470–99.

- Moeschler, Jacques (2013) 'Is a Speaker-Based Pragmatics Possible?', *Journal of Pragmatics* 48(1): 84–97.
- Moore, George (1912). *Ethics*. Oxford University Press.
- Moore, George (1942) 'A Reply to My Critics', in Paul Arthur Schilpp, ed., *The Philosophy of G. E. Moore*: 535–677. Northwestern University Press.
- Moore, George (1944) 'Russell's "Theory of Descriptions"', in Paul Arthur Schilpp, ed., *The Philosophy of Bertrand Russell*: 175–225. Northwestern University Press.
- Murray, Sarah (2021) 'Evidentiality, Modality, and Speech Acts', *Annual Review of Linguistics* 7(12): 1–21.
- Owens, David (2012) 'A Simple Theory of Promising', *The Philosophical Review* 115(1): 51–77. <https://doi.org/10.1215/00318108-2005-002>.
- Pagin, Peter (2016) 'Problems with Norms of Assertion', *Philosophy and Phenomenological Research* 93(1): 178–207. <https://doi.org/10.1111/phpr.12209>.
- Pagin, Peter, and Neri Marsili (2021) 'Assertion', in Edward N. Zalta, ed., *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/archives/win2021/entries/assertion/>.
- Pruss, Alexander R. (2012) 'Sincerely Asserting What You Do Not Believe', *Australasian Journal of Philosophy* 90(3): 37–41.
- Reimer, Marga (1995) 'Performative Utterances: A Reply to Bach and Harnish', *Linguistics and Philosophy* 18(6): 655–75. <https://doi.org/10.1007/BF00983301>.
- Reuter, Kevin, and Peter Brössel (2019) 'No Knowledge Required', *Episteme* 16(3): 303–21. <https://doi.org/10.1017/epi.2018.10>.
- Rosenthal, David (2002) 'Moore's Paradox and Crimmins's Case', *Analysis* 62(2): 167–71.
- Ross, John (1973) 'Slifting', in Maurice Gross, Moris Halle, and Marcel-P. Schützenberger, eds., *The Formal Analysis of Natural Languages*: 133–70. De Gruyter. <https://doi.org/10.1515/9783110885248-009>.
- Rudy Hiller, Fernando (2016) 'Assertion, Justificatory Commitment, and Trust', *Análisis Filosófico* XXXVI(1): 29–53.
- Saul, Jennifer (2012) *Lying, Misleading, and The Role of What Is Said*. Oxford University Press.
- Saul, Jennifer (2017) 'Racial Figleaves, the Shifting Boundaries of the Permissible, and the Rise of Donald Trump', *Philosophical Topics* 45(2): 97–116.
- Saul, Jennifer (forthcoming). *Dogwhistles and Figleaves*. Oxford University Press.
- Sbisà, Marina (2001) 'Illocutionary Force and Degrees of Strength in Language Use', *Journal of Pragmatics* 33: 1791–814.
- Sbisà, Marina (2014) 'Evidentiality and Illocution', *Intercultural Pragmatics* 11(3): 463–83. <https://doi.org/10.1515/ip-2014-0021>.
- Sbisà, Marina (2018) 'Varieties of Speech Act Norms', in Maciej Witek and Iwona Witczak-Plisiecka, eds., *Normativity and Variety of Speech Actions*: 23–50. Brill. https://doi.org/10.1163/9789004366527_003.
- Scanlon, Thomas (1990) 'Promises and Practices', *Philosophy & Public Affairs* 19(3): 199–226.

- Schechter, Joshua (2017) ‘No Need for Excuses: Against Knowledge-First Epistemology and the Knowledge Norm of Assertion’, in Adam Carter, Emma C. Gordon, and Benjamin Jarvis, eds., *Knowledge First*: 132–59. Oxford University Press. <https://doi.org/10.1093/oso/9780198716310.003.0007>.
- Searle, John (1969) *Speech Acts*. Cambridge University Press.
- Searle, John (1989) ‘How Performatives Work’, *Linguistics and Philosophy*, 12(5): 535–58. <https://doi.org/10.2307/25001359>.
- Searle, John, and Daniel Vanderveken (1985) *Foundations of Illocutionary Logic*. Cambridge University Press.
- Shaffer, Michael J. (2012) ‘Moorean Sentences and the Norm of Assertion’, *Logos & Episteme* 3(4): 653–58. <https://doi.org/10.5840/logos-episteme20123414>.
- Siebel, Mark (2003) ‘Illocutionary Acts and Attitude Expression’, *Linguistics and Philosophy* 26(3): 351–66. <https://doi.org/10.1215/00318108-111-1-152>.
- Siebel, Mark (2020) ‘The Belief View of Assertion’, in Sanford Goldberg, ed., *The Oxford Handbook of Assertion*: 97–118. Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190675233.013.4>.
- Simons, Mandy (2007) ‘Observations on Embedding Verbs, Evidentiality, and Presupposition’, *Lingua* 117(6): 1034–56. <https://doi.org/10.1016/j.lingua.2006.05.006>.
- Slote, Michael (1979) ‘Assertion and Belief’, in Jonathan Dancy, ed., *Papers on Language and Logic*: 177–91. Keele University Library.
- Stokke, Andreas (2017) ‘Proposing, Pretending, and Propriety’, *Australasian Journal of Philosophy* 95(1): 178–83. <https://doi.org/10.1080/00048402.2016.1185739>.
- Timmermann, Felix, and Emanuel Viebahn (2020) ‘To Lie or to Mislead?’, *Philosophical Studies*, 178: 1481–501. <https://doi.org/10.1007/s11098-020-01492-1>.
- Unger, Peter (1975) *Ignorance*. Clarendon Press.
- Urmson, James O. (1952) ‘Parenthetical Verbs’, *Mind* 61(244): 480–96.
- Vanderveken, Daniel (1980) ‘Illocutionary Logic and Self-Defeating Speech Acts’, in John Searle, Ferenc Kiefer, and Manfred Bierwisch, eds., *Speech Act Theory and Pragmatics*: 247–72. Springer. https://doi.org/10.1007/978-94-009-8964-1_12.
- Viebahn, Emanuel (2021) ‘The Lying/Misleading Distinction’, *Journal of Philosophy* 118(6): 289–319.
- Vlach, Frank (1981) ‘Speaker’s Meaning’, *Linguistics and Philosophy* 4: 359–91.
- Weiner, Matthew (2005) ‘Must We Know What We Say?’, *The Philosophical Review* 114(2): 227–51. <https://doi.org/10.1215/00318108-114-2-227>.
- Whiting, Daniel (2012) ‘Stick to the Facts: On the Norms of Assertion’, *Erkenntnis* 78(4): 847–67. <https://doi.org/10.1007/s10670-012-9383-6>.
- Williams, John (1982) ‘The Absurdities of Moore’s Paradoxes’, *Theoria* 48(1): 38–46. <https://doi.org/10.1111/j.1755-2567.1982.tb00481.x>.
- Williams, John (2015) ‘Moore’s Paradox in Speech: A Critical Survey’, *Philosophy Compass* 10(1): 10–23. <https://doi.org/10.1111/phc3.12187>.
- Williamson, Timothy (2000) *Knowledge and Its Limits*. Oxford University Press.

Woods, Jack (2018) 'A Commitment-Theoretic Account of Moore's Paradox', in Ken Turner and Laurence Horn, eds., *Pragmatics, Truth and Underspecification*: 323–48. Brill. <https://doi.org/10.1163/9789004365445>.