On Deniability

Alexander Dinges & Julia Zakkou

forthcoming in *Mind*

Abstract

Communication can be risky. Like other kinds of actions, it comes with potential costs. For instance, an utterance can be embarrassing, offensive, or downright illegal. In the face of such risks, speakers tend to act strategically and seek 'plausible deniability'. In this paper, we propose an account of the notion of deniability at issue. On our account, deniability is an epistemic phenomenon. A speaker has deniability if she can make it *epistemically* irrational for her audience to *reason* in certain ways. To avoid predictable confusion, we distinguish deniability from a practical correlate we call 'untouchability'. Roughly, a speaker has untouchability if she can make it *practically* irrational for her audience to *act* in certain ways. These accounts shed light on the nature of strategic speech and suggest countermeasures against strategic speech.

1 Introduction

Communication can be risky. Like other kinds of actions, it comes with potential costs. For instance, an utterance can be embarrassing, offensive, or downright illegal. In the face of such risks, speakers tend to act strategically: they seek 'plausible deniability' (Lee and Pinker, 2010, p. 785) and try to 'minimize their conversational commitments' (Camp, 2018, p. 41). Indirect forms of communication such as insinuation or innuendo, or the use of code words or dogwhistles, are vividly discussed means to accomplish these goals.[1]

While retaining deniability is usually beneficial for the speaker, it often poses problems for the hearer. We want to hold speakers accountable for their conversational behavior, but when speakers retain deniability, they leave us in a 'frustrating position, insofar as a commitment [we] want to reject has been thrust into the conversation, but in a way that escapes easy response' (Camp, 2018, p. 46). How do we avoid this frustrating position?

This paper offers an account of deniability that furthers our understanding of strategic speech and enables us to develop countermeasures against strategic speech. We propose that deniability is an epistemic notion. A speaker has deniability if she can make it *epistemically* irrational for her audience to *reason* in certain ways. To address cases of so-called 'implausible deniability' (e.g. Camp, 2018 and Berstler, 2019), we distinguish this epistemic notion from a practical correlate we call 'untouchability'. Roughly, a speaker has untouchability if she can make it *practically* irrational for her audience to *act* in certain ways.[2]

---

[1] See e.g. Walton, 1996; Pinker, Nowak, and Lee, 2008; Lee and Pinker, 2010; Fricker, 2012, p. 25; Camp, 2018, p. 44 on insinuation and e.g. Jamieson, 1992, pp. 84–93; Mendelberg, 2001; Stanley, 2015, p. 156; Khoo, 2017, p. 47; Saul, 2018, p. 365 on code words.

[2] On our account, both deniability and untouchability come out as distinct from the familiar notion of cancellability; see footnotes 22 and 24.

Our paper is structured as follows. We begin with sample cases in which speakers act strategically by seeking deniability. On this basis, we specify the theoretical role that the notion of deniability is supposed to play (§2). We uncover what one might call the logical form of deniability statements (§3) and present desiderata on adequate accounts of deniability (§4). We briefly discuss accounts of deniability that we find wanting (§5) and then offer our own account (§6). We suggest a generalization of our account (§7) before we address cases of implausible deniability as putative counter-examples (§8). We end with conclusions on strategic speech and on how to limit the scope of deniability we grant to our interlocutors (§9).

2   The theoretical role of deniability

The term 'deniability' is at least semi-technical, and so we cannot exclusively rely on ordinary intuitions to constrain accounts of deniability. In this section, we thus outline the theoretical role that the notion of deniability is supposed to play. This theoretical role will serve to generate desiderata for satisfactory accounts of deniability.

Let us begin with four sample cases taken from the literature where speakers act strategically by seeking deniability. We stipulate that, in each of these cases, the speaker is lying in that she has the communicative intention she is denying. We further assume that everything else is normal (e.g. people can hear each other, they share a common language, etc.). In the first two cases, the speakers offer a bribe through insinuation.

Speeding Driver: *A driver stopped for speeding, to the police officer.*

  'I'm in a bit of a hurry. Is there any way we can settle this right now?'

Denial: 'I didn't mean to offer a bribe! (I was just wondering if I could pay my fine right away.)' (See Lee and Pinker, 2010, p. 790.)

Restaurant: *A customer waiting to be seated, to the waiter.*

  'Is there any way to shorten my wait?'

Denial: 'I didn't mean to offer a bribe! (I was just wondering if you could e.g. split us up on separate tables.)' (See Lee and Pinker, 2010, p. 794.)

In the next two cases, the speakers are making implicit suggestions or statements through insinuation or the use of 'inner city' as a code word.

Realtor: *A realtor, to a potential buyer from a different racial background.*

  'Perhaps you would feel more comfortable locating in a more . . . transitional neighborhood, like Ashwood?'

Denial: 'I didn't mean to convey that you don't fit here. I only meant that Ashwood is on an upward trajectory and might be more interesting to you for this reason.' (See Camp, 2018, p. 43.)

Inner City: *A politician in a press briefing.*

  'I oppose the food stamp program because it primarily benefits inner-city Americans'

Denial: 'I didn't mean to convey anything racist. I just think that the urban poor already receive enough governmental assistance.' (See Khoo, 2017, p. 40.)

2

In all four cases the speakers could have conveyed their messages directly instead of insinuating or using code words, by asking e.g. 'Can we settle this with a bribe?'. Why didn't they?

We run with the following assumption, which is both plausible and widely accepted.[3] The speakers speak the way they speak at least partly because they want to retain deniability regarding the claim that they meant to bribe their interlocutor or that they meant to suggest something racist. In what follows, we develop an account of deniability that allows deniability to play this theoretical role. By accommodating this theoretical role, we will be able to accommodate another, closely related theoretical role. The speakers in the above cases not only insinuate or use code words because they *want* deniability. It seems that they go on to deny their communicative intentions because they think they *have* deniability. At least, it seems that they would not deny their communicative intentions if they did not think they had deniability. We will leave this further theoretical role implicit for now, but we will come back to it when we present our account.

Notice a couple of assumptions that we do not make. First, we do not assume that a desire to retain deniability is the only reason one can have for speaking indirectly (see e.g. Camp, 2018, p. 44 for a brief overview of other reasons). Likewise, we do not assume that denials of one's communicative intention must be based on the belief that one has deniability (one may e.g. issue a denial only to rile up one's audience even further). We merely assume that deniability is a key factor in cases like the ones above. Second, we do not assume that the direct-indirect distinction maps onto the deniable-undeniable distinction. Indeed, our analysis will reveal that directly communicated contents are sometimes deniable while indirectly communicated contents are sometimes undeniable. Finally, we do not assume that the speakers in the above cases have deniability. For now, we only assume that they think they have it and are driven by a desire to retain it.

Notice also that, besides the indicated theoretical roles, there may be further theoretical roles for the notion of deniability to play. For instance, van Elswyk (2020, p. 1168) suggests to use deniability as a criterion to distinguish lying from misleading, Berstler (2019, p. 27) and Pepp (2018, p. 842) float competing ideas on how deniability grounds moral differences between lying and misleading, Reins and Wiegmann (2021) use deniability as a measure of commitment, and according to Saul (2017, p. 109) figleaves can be used to provide deniability. We are optimistic that our notion of deniability can play at least some of these further theoretical roles but we leave a thorough discussion for another occasion.[4]

3   Logical form

This section uncovers what one might call the logical form of deniability statements. We suggest that deniability is relative to what one can deny, to a context and to a part of the audience. Accounts of deniability should respect this structural constraint.

First, deniability is relative to what one can deny. The speeding driver, for instance, could say 'I didn't mean to offer a bribe' and if this denial sticks (in a way to be specified), she has deniability relative to what she denies, i.e., the proposition that she meant to offer a bribe. She might still lack deniability relative to, say, the proposition that she meant to ask

---

[3]See footnote 1 for references.

[4]For a notion of deniability that is distinct from ours, see e.g. Viebahn, 2020, p. 733. His notion is about 'sincere denials' while in cases of deniability in our sense, the relevant denials are usually insincere.

whether the police officer is willing to settle things right now. For simplicity, we will assume that all denials are denials of propositions, and thus we get the following preliminary form.

LF1   S has deniability relative to the proposition p.

Simplifying further, we assume that deniability only applies to propositions about one's communicative intentions, as represented in LF2, where φ stands for the description of a speech act, broadly construed, such as offering a bribe, asking a question, insulting someone, conveying something, etc.

LF2   S has deniability relative to the proposition that she meant to φ.

Deniability can target other propositions too. For instance, it is possible to have deniability relative to propositions about speech acts rather than communicative intentions, or about further types of actions. Arguably, one can not only deny propositions. Maybe, one can also deny speech acts, or actions more generally, by saying things like 'I deny having suggested that' or 'I deny having done that'. Camp (2018, pp. 44–45) correspondingly suggests that deniability targets combinations M(Q) of a 'proposition, Q, in a mode M: as a contribution of information, a query, a directive, etc.', where these combinations can be seen as representing speech acts. Khoo (2017, p. 38) following Mendelberg (2001) further discusses 'deniable norm violations', where a norm violation seems to be a kind of act but not necessarily a speech act. A complete account of deniability should be able to accommodate such generalizations, and we address this in our discussion of accounts of deniability below.

Second, deniability is relative to a context. Speakers may e.g. lack deniability in their current, actual context while retaining deniability in some other context.

LF3   S has deniability at context c relative to the proposition that she meant to φ.

Consider the speeding driver. She might have no deniability in her actual conversational context with the police officer. Arguably, though, she would have deniability if she faced a judge in court, should the police officer arrest her. This might be because the judge 'could not make a bribery charge stick in court by the high standard of proof beyond a reasonable doubt' (Lee and Pinker, 2010, p. 790) or because the judge 'lacks full access to the immediate context' (Camp, 2018, p. 50).

Finally, deniability is relative to parts of the audience along the following lines.

LF4   S has deniability at context c relative to the proposition that she meant to φ and
relative to a part a of the audience she has at c.

A politician at a press conference, for instance, has the press as her audience but also the wider public. Within the same context, she may lack deniability relative to the former while having deniability relative to the latter. Depending on one's favored account of deniability, this might be because press correspondents know a lot about the conversational context while the wider public is generally less knowledgeable.

For simplicity, we will mostly leave the context and the audience parameter implicit and work with the condensed form LF2.[5]

---

[5] This section benefited greatly from comments by an editor and a reviewer of this journal.

4 Desiderata

Before we turn to desiderata on accounts of deniability, let us present two background assumptions we will make throughout. First, we assume that deniability is, in some sense or other, a matter of the success of a pertinent denial. The realtor has deniability relative to the proposition that she meant to convey something racist only if she can successfully deny this, where the relevant notion of success remains to be fleshed out.[6] Second, we favor a liberal understanding of the notion of a denial, where an utterance counts as a denial as long as the speaker denies something by making the utterance. For concreteness let us focus on denials of the proposition that the speaker meant to φ. Such denials need not be simple, explicit denials like 'I didn't mean to φ'. They may be accompanied by an alternative interpretation as in 'I didn't mean to φ. I only meant to ψ' (see e.g. Camp, 2018, p. 45), and they may be implicit as in 'You heard me wrong' or 'I misspoke' (see e.g. García-Carpintero, 2018, pp. 203–204, Berstler, 2019, p. 28 and Boogaart, Jansen, and van Leeuwen, 2020). In the latter cases, the speaker denies having said something, or having intended to say something, and only thereby denies that she meant to φ. One could invoke narrower conceptions of denials, and we will come back to this at the end of this section.

Based on these considerations, this section presents desiderata on accounts of deniability. We start with three key desiderata. They result from the previously discussed idea that a desire to retain deniability explains why the speakers in the above example cases— henceforth, the *target cases*—choose to speak indirectly in the way they do.

A speaker's desire to retain deniability can explain why she chooses to speak a certain way only if the speaker actually has this desire. And she presumably has this desire only if deniability is desir*able*, i.e., only if deniability has some kind of value to the speaker that explains why it is desired. This gives us the first desideratum on adequate accounts of deniability.

Desideratum 1   In the target cases, deniability is desirable for the speaker.

The desire to retain deniability by itself does not explain why speakers speak indirectly rather than directly unless speaking indirectly is a means to fulfil this desire while speaking directly is not. This gives us the following desiderata.

Desideratum 2   In the target cases, speaking directly isn't a means to retain deniability.[7]

Desideratum 3   In the target cases, speaking indirectly is a means to retain deniability.[8]

Respecting the previous relativizations, we can render these desiderata more precisely as follows:

There is some speech act φ, context c, and part a of the speaker's audience at c such that:

---

[6] Lee and Pinker (2010, p. 791) and Camp (2018, p. 45) sometimes seem to suggest that deniability entails a prior strategic intention to deny the target message under certain circumstances. While this may be true for some notion of deniability, it does not hold for ours (see also Desideratum 2 and its discussion).

[7] More precisely: In the target cases, directly performing the respective indirectly performed speech act isn't a means to retain deniability.

[8] More precisely: In the target cases, speaking in the way the respective speaker speaks (i.e. using the respective form of indirect communication in the respective context) is a means to retain deniability.

Desideratum 1′   In the target cases, deniability relative to ϕ, c, and a is desirable for the speaker.

Desideratum 2′   In the target cases, speaking directly isn't a means to retain deniability relative to ϕ, c, and a.

Desideratum 3′   In the target cases, speaking indirectly is a means to retain deniability relative to ϕ, c, and a.

To illustrate, the speeding driver might be aware that she cannot retain deniability in her present conversation with the police officer and so she might aim only at deniability in a court of law. This suffices to explain her choice to speak indirectly if indirect but not direct speech allows her to reach this goal. Lee and Pinker (2010, p. 796) similarly suggest that indirect communication may be motivated already when a denial is 'plausible *to a virtual audience*, even if it is plausible to neither the speaker nor the hearer'. We will consider Desiderata 1′ to 3′ when precision matters, but for the most part, we will work with the simplified Desiderata 1 to 3.

Let us turn to secondary desiderata on accounts of deniability. We start with two desiderata that specify different ways in which deniability is consistent with the audience being in a strong epistemic position.

First, think of Grice's (1989) familiar petrol case, where A tells B, who has run out of gas, 'There is a garage around the corner'. Initially, A seems to have deniability vis-à-vis the proposition that she meant to convey that the garage might sell petrol. If A issued a denial right away, by saying 'I didn't mean to suggest that you might get petrol there. I only meant that you could get a newspaper to pass the time', B would (and should) believe her and grant that A did not have the respective communicative intention. Meanwhile, if A does not issue the denial, B knows that A meant to convey that the garage might sell petrol.[9] It thus seems that a speaker can have deniability about whether she meant to ϕ even though her audience knows that she meant to ϕ. This gives us the following desideratum.

Desideratum 4   Deniability is consistent with the audience knowing the speaker's communicative intentions, in cases where the respective denial is not actually made.

To support this desideratum further, notice that many people hold that deniability is widespread, being a relatively general feature of indirect communication and maybe even of some types of direct communication involving context-sensitive expressions (e.g. Fricker, 2012 and Peet, 2015). Unless we accept the indicated desideratum, this implies radically skeptical consequences for testimonial knowledge. After all, if a hearer does not know that a speaker meant that p, it is hard to see how she can come to know that p based on the speaker's testimony (see also Davies, 2019, pp. 23–24).[10]

---

[9]One may question this verdict depending on one's favored account of knowledge. What ultimately matters to us is that B meets whatever epistemic standard suffices for actionability in the context at hand, be it knowledge or something else. See §7.

[10]One might object that testimonial knowledge does not require knowing *exactly* what the speaker meant. To avoid such concerns, one can just restate Desideratum 4 in terms of knowledge of what the speaker *roughly* meant and adjust the rest accordingly. One may object that testimonial knowledge does not even require such rough knowledge (e.g. Peet, 2018). We cannot address this thesis here, except to note that it is a minority position.

Second, 'deniability' and 'plausible deniability' are sometimes used interchangeably. However, almost everybody in the debate agrees that even if a speaker has deniability, her denials need not be plausible. Many authors hold, for instance, that the speakers in the target cases from above have deniability despite the fact that the relevant denial would be quite implausible (e.g. Lee and Pinker, 2010, p. 793; Camp, 2018, p. 48; Berstler, 2019, pp. 27–28). We get the following desideratum.

Desideratum 5   Deniability is consistent with the audience finding the denial implausible, in cases where the respective denial is actually made.

By granting this desideratum, we acknowledge that there are cases of 'implausible deniability' in one sense of the term. Usually, though, cases of implausible deniability are construed as cases where the respective denial is not just implausible but mutually known to be false. For now, we stay neutral on whether there are cases of implausible deniability in this latter sense, but we challenge this assumption in §8.

The desiderata we present next specify factors on which deniability can depend. Whether a speaker has deniability can depend on the whole range of factors relevant for Gricean calculability, such as the sentence's conventional meaning, background knowledge and expectations about cooperativeness (Grice, 1989). This is obvious enough. The two factors we present below are noteworthy because they usually have no place in Gricean calculations.

First, Peet (2015, p. 48) suggests that 'repeat offenders' who continuously mislead about their communicative intentions may lose deniability. They presumably lose deniability because their audiences no longer trust them when they deny their communicative intentions. This gives us the following desideratum.

Desideratum 6   Whether a speaker has deniability can vary with how trustworthy her denials are.

Three clarificatory remarks. First, we can trust speakers in one domain but not another. For instance, we trust the weather-caster when it comes to the weather but not when it comes to politics. On the above desideratum, deniability depends on trust *concerning the respective denials*. It need not depend on the general trustworthiness of the speaker or her trustworthiness in other domains.[11] Second, the above desideratum entails that a loss of trust can lead to a loss of deniability. It does not entail, however, that deniability requires even a minimum amount of trust. In some contexts, you may have deniability while being completely untrustworthy, say, because your communicative intentions are obscure independently of whether you deny them. Third, the above desideratum only entails that one's trustworthiness *can* affect the scope of one's deniability. To establish this desideratum, we need just one case where a speaker lacks deniability but would have deniability if her denials were more trustworthy. Peet's repeat offender seems to be a case in point.

Next, suppose a speaker deliberately mumbles 'Hannah is an idiot', and when challenged, she responds 'You heard me wrong. I said that Hannah is a chariot. She's unstoppable'. In a case like this, it seems natural to say that the speaker secures deniability by

---

[11]Desideratum 6 is thus consistent with Camp's (2018, p. 51) observation that 'a known likelihood of conflict or other motivation for strategic interpretation [...] widens the scope of deniability' even though an increased 'likelihood of conflict or other motivation for strategic interpretation' presumably leads to lesser trust overall. Lesser trust *overall* can widen the scope of deniability even if lesser trust *regarding denials* narrows down the scope of deniability.

mumbling. We can likewise retain deniability by e.g. choosing a noisy environment where it is difficult to understand each other. We get the following desideratum.

Desideratum 7   Whether a speaker has deniability can vary with presentational factors (e.g. mumbling) and environmental factors (e.g. noise) bearing on the probability of misspeaking and mishearing.

To be sure, the relevant denials here are denials like 'You heard me wrong' rather than 'I didn't mean to φ'. The present desideratum therefore requires our indicated wide conception of denials. One could invoke narrower conceptions of denials restricting one's attention to e.g. simple, direct denials. But with the mumbling case in mind, this would implausibly and unnecessarily limit the applicability of the notion of deniability. Again, it seems natural to say that Hannah mumbles because she wants to retain deniability, and one should not foreclose this explanatory role for the notion of deniability from the start.

This concludes our collection of desiderata. We now turn to the question of how to accommodate them, starting with accounts of deniability that we find wanting.

5   Accounts

Consider the following definition of deniability first (see e.g. Mazzarella et al., 2018, p. 16).

Deniability 1   S has deniability relative to the proposition that she meant to φ iff S can deny that she meant to φ.

This definition is problematic unless we specify the relevant sense of 'can'. On the 'can' of ability, for instance, it would credit everybody who is able to communicate with universal deniability. After all, competent communicators are generally able to deny their communicative intentions by e.g. producing utterances of the form 'I didn't mean to φ.' This is at odds with Desideratum 2, according to which speaking directly is not a means to retain deniability in the target cases.

Consider the following restriction of Deniability 1.

Deniability 2   S has deniability relative to the proposition that she meant to φ iff S can plausibly deny that she meant to φ.

This definition is untenable as well because it directly clashes with Desideratum 5, according to which deniability does not entail plausibility.

Consider the following alternative restriction of Deniability 1.

Deniability 3   S has deniability relative to the proposition that she meant to φ iff S can properly deny that she meant to φ.

This definition is problematic for the following reason. Lies are improper because they violate the norm of assertion. According to the knowledge norm of assertion, for instance, one can properly assert p only if one knows p (Williamson, 2000), and in cases of lying, this condition is blatantly violated. Now, in the target cases, the denials in question are lies because the speaker has the respective communicative intentions (by assumption). Consequently, these denials are improper and the speakers lack deniability on the given definition,

independently of whether they speak indirectly. This is at odds with Desideratum 3, according to which speaking indirectly should be a means to retain deniability. Analogous considerations hold for familiar alternatives to the knowledge norm of assertion, such as justified belief norms (e.g. Lackey, 2007) or truth norms (e.g. Weiner, 2005).[12]

Let us turn to a more elaborate and more promising account of deniability, which is extracted from Camp (2018) and adapted to our terminology. Peet (2015, p. 31), following Fricker (2012), offers a similar definition (see also Davies, 2019, p. 19).

Deniability 4   S has deniability relative to the proposition that she meant to ɸ iff it is '*reasonable* to calculate [that the speaker meant to ψ (≠ ɸ)] on the basis of the uttered sentence's conventional meaning [. . .], the commitments undertaken in the conversation to this point, and some set [π'] of epistemically accessible presuppositions consistent with those commitments, in a way that renders [S's utterance] at least minimally conversationally cooperative.' (Camp, 2018, p. 50)

Very roughly, a speaker has deniability on this account when their utterance is open to more than one interpretation. This definition makes progress because the previously problematic desiderata now seem to be satisfied. Indirect as opposed to direct communication can open up new interpretative options, but these interpretative options need not be particularly plausible.

Still, the definition is problematic. First, it doesn't generalize to denials of actions other than speech acts (likewise for propositions about them). Following e.g. Khoo (2017, p. 38), we assumed that norm violations can be deniable and thus that non-communicative actions can be deniable. For instance, a committee might deniably dismiss scholarship applications on racist grounds. According to Deniability 4, this would mean that we can somehow calculate e.g. that the committee's decisions were based on merit alone. However, it seems impossible to calculate anything like that. At least in a Gricean framework, the relevant calculations always terminate with claims about a speaker's communicative intentions, not with claims about non-communicative actions (e.g. Grice, 1989, p. 31). To accommodate deniable norm violations, one would have to broaden one's understanding of calculability accordingly. Maybe this can be done, but we are skeptical.

Second, Deniability 4 is at odds with Desideratum 6, according which deniability can vary with the trust we place in the speaker's denials, and Desideratum 7, according to which factors bearing on misspeaking and mishearing affect deniability. For whether a speaker's denials are trustworthy relevantly affects neither the 'epistemically accessible' presuppositions nor what one can 'reasonably' calculate on their basis, and whether a speaker e.g. mumbles or whether it is e.g. noisy does not relevantly affect these factors either. For instance, the question under discussion and background assumptions about how fines are normally payed might make it reasonable to calculate that the speeding driver only meant to ask whether she can pay their fine right now. A loud truck rushing by, however, does not.

---

[12]One could propose weaker understandings of propriety to respond (e.g. in terms of 'felicity'), but such notions would have to be spelled out, while bearing in mind Desideratum 2. Berstler (2019, p. 28), e.g., can be taken to suggest a comparative notion of propriety, where deniability entails that the respective denial 'sounds better' than a straightforward denial of a literally expressed communicative intention. As she notices herself, however, this would make deniability universally available, through utterances of e.g. 'I misspoke, I meant to ψ', contrary to Desideratum 2. Of course, Berstler might have a different theoretical role for her notion of deniability in mind than we have.

A final problem takes the form of a dilemma. According to Deniability 4, deniability means that the speaker's utterance is, in a specific way, open to more than one interpretation. Does this entail that the hearer does not know what the speaker meant? If so, Deniability 4 clashes with Desideratum 4, according to which hearers may have such knowledge despite deniability. If not (e.g. because the alternative interpretations do not count as 'relevant alternatives' (Dretske, 1970)), it becomes difficult to explain Desideratum 1, according to which deniability is desirable. For if deniability is consistent with the audience *knowing* one's communicative intentions, then why would deniability be desirable? If the audience knows a realtor had e.g. racist intentions, they can blame her for this, they can tell others about this, and they can generally let their actions be guided by this. So what does the realtor gain from retaining deniability? Once more, this is a challenge rather than an objection, but together with the previous concerns, we take it as sufficient motivation to look for an alternative proposal.

6   Our proposal

We have proposed desiderata on accounts of deniability, and we have shown that it is difficult to meet them. It is time to turn to our own favored analysis. We begin with a simplified proposal. We use this proposal to illustrate some basic features of our official account, which we present in the next section.

Deniability   S has deniability relative to the proposition that she meant to ϕ iff: if S denies that she meant to ϕ, then S's audience does not know that she meant to ϕ.[13,14]

On this account, there are two ways to have deniability. According to the first, you have deniability because your audience does not know what you meant whether or not you issue a denial. This may be for various reasons, and the use of figleaves in prior discourse may be noteworthy among them (see e.g. Saul, 2017).[15] According to the second, you have deniability because your denial undermines otherwise present knowledge. In paradigmatic cases of deniability, a denial undermines knowledge because it provides misleading evidence about the speaker's communicative intentions such that the resulting total body of evidence no longer suffices for knowledge.[16] But denials can undermine one's knowledge in other ways too. For instance, an irrationally gullible audience might lose their beliefs about the speaker's communicative intentions and thereby their knowledge despite the fact that these beliefs would amount to knowledge (we assume that knowledge entails belief; see e.g. Myers-Schulz and Schwitzgebel, 2013; Rose and Schaffer, 2013 for discussion). Or suppose that as soon as you issue a denial, a friend backs you up ('I know him. He

---

[13]Our proposal is congenial to Berstler's (2019, p. 27), who suggests that 'I have plausible deniability [when] my interlocutors won't know whether my claim "I didn't mean p" is true or not'. But, first, this definition deserves much deeper discussion than Berstler aims to provide. Second, it must be modified in certain respects; see §7. Third, we think it should be extended to deniability full stop, whether plausible or not; see §8.

[14]The conditional in our definiens is an indicative conditional. Those who treat it as equivalent to a material conditional might opt for a subjunctive conditional instead. However, any presupposition or implicature of counterfactuality should be detached. One can have deniability even if one has denied ones communicative intentions already. Thanks to an editor for urging us to clarify this.

[15]Thanks to Nikki Ernst for drawing our attention to figleaves.

[16]This suggests that knowledge is defeasible, contra e.g. Lasonen-Aarnio, 2010. We can dispense with this assumption once we get to our official definition in §7. See footnote 26.

would never suggest anything like that!'). Your audience might lose knowledge only due to the collective evidential weight of your denial together with your friend's support. In all these cases, the speaker has deniability on our definition because the respective conditional is true.[17]

While our account is officially stated in terms of propositions about communicative intentions, we can easily extend it to propositions about speech acts and other actions. We just have to rephrase it in terms of propositions p rather than propositions that the speaker meant to φ. We can also accommodate the deniability of actions (rather than propositions about them) once we assume that an action is deniable iff the proposition that one performed this action is deniable.[18]

Our secondary desiderata are easily accommodated too. Desideratum 4 ('knowledge'): On the given definition, deniability only entails that the audience lacks knowledge *if* the speaker issues a denial. If she does not, the audience may know what was meant. Desideratum 5 ('plausibility'): On the given definition, the audience does not know a respective denial is false in cases of deniability. They might still find the denial implausible. After all, we consider many propositions implausible even though we do not know that they are false. Desideratum 6 ('trust'): If you are trustworthy, your denials have more evidential weight and are therefore more likely to undermine your audience's knowledge. This makes it more likely that you have deniability on the given definition. Desideratum 7 ('misspeaking and mishearing'): If you e.g. mumble or speak in a noisy environment, it is more likely that your audience does not know what you said and therefore what you meant. Again, this makes it more likely that you have deniability on our account.

Let us now turn to our key desiderata, which deserve an extended discussion. We begin with Desideratum 1, according to which deniability is desirable in the target cases. Many people assume a knowledge norm of action along the following lines (e.g. Hawthorne and Stanley, 2008).

KN-A It is epistemically proper to treat p as a reason for action iff one knows p.

For instance, you can properly treat the proposition that your friend will be at the party as a reason to go to the party only if you know that your friend will be at the party. Given KN-A and Deniability, it is clear why deniability is often desirable. If you have deniability, you can rob your audience of knowledge and therefore of reasons for adversarial actions against you. For instance, the realtor can rob the buyer of reasons to blame her for being racist.

Naturally, emotional reactions such as resentment and anger are governed by reasons too.[19] Friends of KN-A should thus be sympathetic to the following principle.

---

[17]To us, it seems natural to classify cases like the gullibility and the support case as featuring deniability (see also our discussion of non-evidential deniability below). Some readers may want to exclude them. They might define deniability in terms of conditionalization rather than conditionals. A speaker has deniability iff the audience's body of evidence does not suffice for knowledge once it is updated with the proposition that a relevant denial has been made.

[18]One may worry about cases where, intuitively, you have deniability while a denial would generate rather than undermine knowledge of your communicative intentions by only raising suspicions (consider e.g. Getrude's 'The lady doth protest too much, methinks'). In response, we could broaden our account such a that a speaker has deniability iff there is some relevant communicative course of action available to them such that if the speaker chooses that course of action, the hearers lack relevant knowledge. Relevant communicative courses of action include issuing a denial but they may also include remaining silent or redirecting your audience's attention. This remains to be fleshed out. Thanks to Helen Han Wei Luo, Per-Erik Milam and Sumeet Patwardhan.

[19]Maguire (2018) opposes this 'dogma', but even he denies only that reasons for affective attitudes are of the same kind as reasons for actions. This seems consistent with what we have to say.

KN-E   It is epistemically proper to treat p as a reason for emotional attitudes iff one knows p.[20]

For instance, you can properly resent someone based on their racist attitudes only if you know they have them. Given KN-E, it is clear why deniability is desirable even in cases where the speaker need not worry about adversarial *actions*. In Restaurant, for instance, the waiter might just politely ignore an attempt to bribe, even if she objects to it. She might still feel resentment, and one may want to retain deniability because one wants to be able to eliminate reasons for such emotional attitudes.[21]

The previous considerations also help to accommodate our second theoretical role for deniability, according to which the speakers in the target cases deny their communicative intentions partly because they think they have deniability (see §2). When you desire something and believe that something is a means to this end, then your desire and your means-ends belief can be reasons for choosing this means. Now, as just discussed, the speakers in the target cases want to undermine the audiences' knowledge. Furthermore, the belief that you have deniability as defined is a belief to the effect that denials are a means to this end. Thus, this latter belief can be a partial reason for issuing a denial, in line with the indicated theoretical role (see also our discussion of Desideratum 3).

Let us turn to Desideratum 2, according to which speaking directly is not a means to retain deniability in the target cases. We can straightforwardly accommodate this desideratum as follows. If e.g. the speeding driver were to directly offer a bribe, her audience would know she offered a bribe even if she denied this. Thus, she would not retain deniability by speaking directly (holding everything else fixed). Turning to the more precise Desideratum 2′, she would not even retain deniability in relevant counterfactual context. For instance, even a judge would know what she meant based on the police officer's testimony.

Note that our account still does not imply that speaking directly is never a means to retain deniability. Think of cases where it is easy to mishear one another. Here, even literally expressed contents can be deniable because the hearers can fail to know what was said. Or think of situations in which we are prone to misspeak or where we can convincingly appeal to irony or metaphor, or where we can convincingly 'feign ignorance' (Stanley, 2015, p. 156) of the literal meaning of a given word or sentence. Here, too, even semantically expressed contents can be deniable, not because the hearer does not know what the speaker said but because she does not know whether the speaker meant what she said.[22]

Consider Desideratum 3, according to which speaking indirectly is a means to retain deniability in the target cases. This desideratum may seem problematic for our account. It may seem that, in Speeding Driver for instance, the police officer knows that the speeding

---

[20]Buchak (2014, p. 299), e.g., considers a knowledge norm for blaming.

[21]Lee and Pinker (2010, p. 795) similarly suggest that speakers sometimes speak indirectly to avoid 'the risk of awkwardness or shame in the same way that a briber avoids the risk of an arrest.' But they focus only on such emotions as a result of a clash between 'the relational model assumed by the speech act' and 'the model that currently holds between the speaker and hearer'. This seems too narrow. When someone has racist communicative intentions, their worry is not that they shift the 'relational model' in some unappreciated way, but rather that they face an audience who condemns racism and thus acquires a reason to condemn them.

[22]See Boogaart, Jansen, and van Leeuwen, 2020, pp. 9–16 for an array of real-life cases of attempted denials of literal contents (see also Michaelson, 2021 on 'sneaky reference'). These considerations suggest that deniability does not entail cancelability (*pace* Lee and Pinker, 2010, p. 791 and Mazzarella et al., 2018, p. 16). After all, the literal meaning of a sentence is usually thought to be non-cancellable. See e.g. Blome-Tillmann, 2008 and Zakkou, 2018 for discussion. Stanley (2015) also suggests that certain not-at-issue contents are deniable (156) but not cancellable (139), but see Saul, 2018, p. 374 for critical discussion of his examples.

driver meant to offer a bribe and that she would retain this knowledge even if the speeding driver denied that she meant to offer a bribe. On our account, it follows that the speeding driver lacks deniability. Given that the speeding driver speaks indirectly, it then follows that speaking indirectly is not a means to retain deniability, contrary to Desideratum 3. We offer three responses to this type of concern, focusing on Speeding Driver for concreteness.

First, it is not obvious that the speeding driver cannot shake the police officer's knowledge with a relevant denial, and so it is not obvious that she lacks deniability. Recall that there are two ways to retain deniability: either your audience does not know what you meant to begin with, or your denial defeats otherwise present knowledge. Now there are two ways to defeat otherwise present knowledge. The first works by undermining the evidential dimension of knowledge and endows you with what we call *evidential deniability*. You have evidential deniability if your audience is no longer *in a position to* know that you meant to φ once you deny that you meant to φ. The second way to defeat otherwise present knowledge works by undermining the psychological dimension of knowledge and endows you with what we dub *psychological deniability*. You have psychological deniability if your audience no longer believes and therefore no longer knows that you meant to φ once you deny that you meant to φ. While psychological and evidential deniability often go together, they can come apart if your audience loses a belief despite the fact that it would amount to knowledge.

Returning to Speeding Driver, we can acknowledge that the speeding driver lacks evidential deniability because her audience's evidence is too strong. The speeding driver may still have psychological deniability. As indicated, speakers can deny their communicative intentions by offering an alternative interpretation of their utterance. The speeding driver, for instance, could say, 'I didn't mean to offer a bribe! I was just wondering if I could pay my fine right away.' Now it is widely held that the salience of such alternatives can affect our thinking in irrational ways. According to Gerken (2017, p. 116), for instance, speakers have a psychological tendency to treat salient possibilities as epistemically relevant even when they are not. Relatedly, Hawthorne (2004, pp. 162–166) and Williamson (2005, p. 226) suggest that salience can make an error-possibility appear more likely than it is (see also Nagel, 2010 and Dinges, 2018b for further discussion). Each of these mechanisms may make the police officer lose her belief about what the speeding driver meant even though her evidence is strong enough for knowledge. Each of these mechanisms may therefore credit the speeding driver with psychological deniability even when she lacks evidential deniability.[23]

Note that our account still does not imply that speaking indirectly guarantees deniability. Some police officers, for instance, might not fall for the indicated psychological effects and if this is so, the speeding driver may lack deniability after all. Or consider the following case from Weiner (2006, p. 128). Someone is sprawled across two seats in a crowded train, and you say, 'I'm curious as to whether it would be physically possible for you to make room for someone else to sit down.' The implicature that the addressee should make room seems undeniable because it is just too obvious that it is there.[24]

---

[23]According to some versions of anti-intellectualism (e.g. Hawthorne, 2004; Dinges, 2018a) and many versions of epistemic contextualism (e.g. Lewis, 1996; DeRose, 2009; Blome-Tillmann, 2014), salient error-possibilities may destroy knowledge or the ascribability of 'knowledge' because they tend to raise the epistemic standard. This yields another form of non-evidential deniability potentially available to the speeding driver.

[24]This suggests that cancelability does not entail deniability. For on many accounts of cancelability (e.g. Blome-Tillmann, 2008 and Zakkou, 2018), the indicated implicature is still cancellable. Weiner (2006) suggests an account of cancelability where this is questionable, but we think he confuses cancelability with deniability.

Second, recall that, officially, our third desideratum is Desideratum 3′ rather than Desideratum 3. Even if we cannot accommodate the latter desideratum because the speaker lacks deniability, we may still be able to accommodate the former. This desideratum is accommodated already if the speeding driver has deniability relative to a relevant part of her audience in some relevant context. Speaking indirectly might not be a means for the speeding driver to retain deniability at her current context, but it might still be a means to retain deniability in court if the police officer decides to press charges (see also Lee and Pinker, 2010, p. 790 and Camp, 2018, p. 50 cited above).[25]

Third, even Desideratum 3′ is a bit too strong. We have assumed so far that speaking indirectly must be an actual means to retain deniability in the target cases. However, to explain why a speaker speaks indirectly, it suffices that she believes, or maybe only hopes, that speaking indirectly is a means to retain deniability. We should weaken Desideratum 3′ accordingly. This opens up further ways to accommodate this desideratum even if we grant that the speeding driver doesn't have deniability in any relevant context relative to any relevant audience. For instance, optimism bias (e.g. Sharot, 2012) may make the speeding driver overly optimistic about the extent of her deniability or she might just be confused about the epistemic standards in court. She might therefore believe she has deniability when in fact she does not.

We think that all target cases, i.e., all cases where a desire to retain deniability supposedly explains indirect speech, fall into one or the other category above. In all these cases, speaking indirectly is a means to retain evidential or psychological, real or perceived deniability in some relevant context relative to some relevant audience. In this way, we can accommodate the basic insight behind Desideratum 3.

Some readers will harbor residual doubts about our discussion of Desideratum 3. Aren't there cases where speakers speak indirectly due to deniability even though every relevant audience knows what they meant even if they deny this? We address such cases of implausible deniability below. First, however, we turn to a generalization of our account.

7 A generalization

The suggested account of deniability is available only to friends of the above knowledge norms, for these norms helped to explain our key Desideratum 1. While knowledge norms are popular, we do not want to rely on them. We therefore suggest the following revised definition.

Deniability′   S has deniability relative to the proposition that she meant to φ iff: if S denies that she meant to φ, then it is not epistemically proper for S's audience to treat this proposition as a reason for actions or emotions.

Given the knowledge norms, this new definition collapses into our previous definition, but if we assume a different norm, these definitions come apart. Our desiderata should still be

---

Lee and Pinker (2010, p. 791) also argue that cancelability does not entail deniability, but their argument relies on a problematic conception of deniability; see footnote 6.

[25] A related hypothesis would be that the speeding driver doesn't aim at deniability in our sense but at what one could call *legal* deniability. A speaker has legal deniability relative to p iff if they deny p, p doesn't count as legally admissible evidence. Legal deniability and deniability in our sense come apart where legal evidence standards do not track knowledge. Thanks to Samia Hesni.

satisfied because the epistemic state that replaces knowledge on alternative norms is usually quite similar to knowledge (see Benton, n.d. for an overview of the discussion).[26]

This definition is still not neutral enough. For the epistemic requirements on reasons may differ for different reactions, say, for emotions as opposed to actions or, within actions, for more or less drastic measures. This would make Deniability$'$ problematic, because one now has to ask *which* reactions matter for deniability. For instance, does the speeding driver have deniability already if the police officer cannot treat the proposition that she meant to offer a bribe as a reason for arresting her? Or does she gain deniability only if the police officer cannot treat this proposition as a reason for resenting her, for scolding her, etc.?

One response would be that one has deniability when one's denial undermines reasons for *some* or for *every* reaction available to the audience. We favor a more flexible approach that further relativizes deniability to a set of reactions. The logical form of deniability and, correspondingly, our key Desiderata 1$'$ to 3$'$ should be adjusted accordingly.

Deniability$''$ S has deniability relative to the proposition that she meant to ϕ and relative to a set of actions or emotions α iff: if S denies that she meant to ϕ, then it is not epistemically proper for S's audience to treat this proposition as a reason for actions or emotions α.

For instance, the epistemic requirements on reasons for resentment may be low, and so the speeding driver may not have deniability when it comes to this reaction. Meanwhile, the epistemic requirements for reasons for arresting someone may be high, and so, by speaking indirectly, the speeding driver may retain deniability relative to this reaction. The latter fact may suffice as a reason to speak indirectly if the speaker cares about not being arrested.

8   Plausible and implausible deniability

Both Camp (2018) and Berstler (2019, pp. 27–28) suggest that there are cases of implausible deniability. In these cases, speakers have deniability and yet the relevant audience knows their communicative intention even when they deny it. Or as Camp (2018, p. 47) puts it, the respective denials are 'bald-faced lies'. Cases of implausible deniability are immediate counterexamples to our initial definition of deniability in terms of knowledge. They also clash with our revised, reasons-based definition if we assume that it is epistemically proper to treat known propositions as a reason; in brief, that knowledge suffices for actionability. For assuming that knowledge suffices for actionability, cases of implausible deniability are cases where a speaker has deniability and yet, even in the face of denials, it is epistemically proper for the audience to treat the target proposition as a reason for relevant reactions. In brief, the argument against our account of deniability rests on the following two premises.

P1   There are cases of implausible deniability.

P2   Knowledge suffices for actionability.

---

[26]The above definition no longer presupposes that knowledge is defeasible; see footnote 16. Only the epistemic status required for properly treating something as a reason comes out as defeasible. One might worry that, given the knowledge norms, it still follows that knowledge is defeasible. We can avoid this conclusion by granting that speakers have deniability already if the respective reasons lack primary or secondary (or tertiary) epistemic propriety. This commits us only to the conclusion that knowledge or the respective secondary (or tertiary) epistemic conditions are defeasible. Proponents of indefeasible knowledge seem happy to accept this; see e.g. Williamson, forthcoming, n. 19.

These premises jointly entail that there are counterexamples to Deniability, Deniability′ and Deniability″. In what follows, we respond to this concern. Our official response denies P1, but we also present ways to deny P2 which are familiar from the debate on the norms of action.[27]

Before we address P1 and the alleged cases of implausible deniability, let us define a practical notion of untouchability alongside the epistemic notion of deniability.

Untouchability   S has untouchability relative to the proposition that she meant to ϕ iff: if S denies that she meant to ϕ, then it is practically irrational for S's audience to engage in pertinent actions or emotions.

Untouchability does not entail deniability as defined. Suppose a speaker had racist communicative intentions and that it is epistemically proper for the audience to treat this intention as a reason for speaking up. The speaker lacks deniability on our definition, but she may still have untouchability. This is because the audience may have reasons that speak against speaking up, and these reasons can make it overall irrational to do so. For instance, speaking up may be socially inappropriate in certain contexts, it may draw more attention to a speaker than she deserves, it may be costly if you are e.g. dealing with a superior or an oppressor, it may be practically difficult if you are e.g. part of a larger crowd, etc.[28]

Now consider supposed cases of implausible deniability. Camp (2018, p. 48) suggests that our target cases can involve implausible deniability. We grant that, on some construals of these cases, the respective denials are implausible in the indicated sense, i.e., they are mutually known to be false. We also grant that, on some (e.g. the typical) construals, the speakers have deniability. However, we doubt that, on any construal, the denials are relevantly implausible *and* the speakers have deniability. Consider Speeding Driver. We have argued at length how, in such a case, the respective denial can undermine the audience's knowledge. Thus, we have shown that the denial in this case is not implausible in the relevant sense. One might think that, on some construals of the case, all our proposals are unconvincing because the speaker's request is too obvious an attempt to bribe. But then one would have to show that the speaker still has deniability. Camp (2018, p. 47) seems to motivate this assumption with the claim that the speaker would 'get away' with their denial. But in Speeding Driver, this is far from obvious. If we envisage the case as one where the speaker e.g. winks and shows the money while speaking, it seems all the more natural that she will not get away with this and that the police officer will react accordingly.[29]

This may be different in Restaurant, Realtor and Inner City. As before, we think that, on many construals of these cases, the audience does not know that the respective denial is false, and thus the denial is not implausible in the relevant sense. This time, though, we grant that, on some construals, the respective denial is relevantly implausible while it would still seem that the speakers 'get away' with it. But this still doesn't entail implausible

---

[27] A reviewer suggests an alternative way to deal with implausible deniability in terms of the following definition of deniability. A speaker has deniability relative to the proposition that she meant to ϕ iff: if she denies this proposition, it is not public. If we tear publicity and common knowledge apart (e.g. Lederman, 2018 and Camp, 2018), this definition would accommodate cases of implausible deniability. However, we would need some other account of publicity that explains e.g. Desideratum 1 ('desirability') and Desideratum 6 ('trust'). It seems worthwhile to explore this further, but we lack the space to do so.

[28] We draw here on a similar list of reasons against speaking up from Goldberg, 2020, p. 177.

[29] The same goes for Berstler's (2019, p. 27) supposed case of implausible deniability. The speaker will not get away with their denial (the expected rejoinder would be 'Don't be ridiculous!'), and thus we see no reason to grant her deniability. See also footnote 12.

deniability, for getting away does not entail deniability. The speakers get away (if they do) because they have untouchability, not deniability. Even if the waiter recognizes the attempt to bribe, debating this may be irrational because she might lose a paying customer. Similarly, even if you know that your realtor is a racist, it may be irrational to challenge their denial if you still want to find a place to live. Calling out a politician likewise comes with the usual costs associated with engaging in public controversies. The distinction between deniability and untouchability undermines the motivation for claiming that the mentioned cases feature deniability. Since there is thus no positive evidence for implausible deniability, we can use P2 together with our account to argue that implausible deniability does not exist.

One could object that what we call untouchability actually is deniability. However, defining deniability as untouchability has odd consequences. A speaker may gain untouchability relative to the proposition that she meant to convey something racist only because someone bribes everybody not to call her a racist. Similarly, a speaker may lose untouchability relative to the proposition that she meant to convey something racist only because someone bribes everybody to groundlessly accuse her of racism. We find it odd to say that speakers gain or lose deniability in this way. Most authors, we included, think of deniability in much more epistemic terms. The above definition Deniability 4 from the literature, for instance, clearly targets a more epistemic notion. To be sure, 'deniability' might have both a practical and an epistemic reading. Practical deniability could then be defined as untouchability. However, our paper should be construed as concerning epistemic deniability only.

These considerations undermine P1, according to which there are cases of implausible deniability. As indicated, we would also like to challenge P2, according to which knowledge suffices for actionability. This premise bears on a wide-ranging discussion of the norms of action and emotion, which we cannot cover here. We'd only like to flag some familiar ways to deny that knowledge always suffices for actionability.

Many authors hold that knowledge suffices for actionability. This follows from the previous knowledge norms, and it also follows from norms that appeal to e.g. truth or justified belief. After all, knowledge entails true, justified belief. Gerken (2017, pp. 130–150), however, defends a context-sensitive alternative to such norms where the epistemic requirements on reasons sometimes rise above the standard of knowledge (see also Brown, 2008; Reed, 2010). Simplified and adjusted to the present terminology, he suggests the following principle.

WN-A It is epistemically proper to treat p as a reason for action iff one is warranted in believing p to a degree that is adequate relative to one's present deliberative context.

Here, 'warrant' is used as 'a broad label for epistemic rationality' (Gerken, 2017, p. 10), while the following is a non-exhaustive list of factors that determine the 'deliberative context': 'alternative *courses of action*, availability of *further evidence*, considerations of *urgency*, the *stakes* associated with the action, the *social roles* and *conventions* associated with the action' (133). Depending on these factors, the warrant required for actionability may go beyond knowledge according to Gerken (141–143).

Importantly, knowledge and actionability may come apart specifically in putative cases of implausible deniability. Suppose a strategic speaker insinuates racist messages. The stakes for calling her out might be high due to the high social costs associated with being a

racist.[30] This might raise the epistemic standards beyond knowledge. At least, the *perceived* epistemic standard may be very high. Similarly, police officers and judges, say, occupy specific social roles with associated conventions. These conventions might place specific epistemic burdens on them that might go beyond knowledge.

Some proponents of implausible deniability might be sympathetic to a framework like Gerken's. Camp (2018, pp. 51–52), for instance, seems to suggest that deniability in the context of sexual harassment is constrained by an operative 'reasonable person' standard in some legal and administrative domains. Lee and Pinker (2010, p. 790) relatedly appeal to the 'high standard of proof beyond a reasonable doubt' in court to explain deniability. Gerken's framework offers a straightforward explanation for how such legal and administrative standards can have a bearing on the scope of a speaker's deniability. By establishing such standards, we change the conventions associated with relevant reactions towards e.g. sexual harassment. Such conventions are part of the deliberative context. Thus, on Gerken's view, they set the epistemic standard for relevant reactions and, on our definition, the epistemic standard for deniability.[31]

In sum, we are skeptical that cases of implausible deniability exist. But even if they do, it does not follow that our definition fails. One still has to establish that knowledge suffices for actionability in the relevant cases, which is far from obvious.

9 Upshots

In this section, we present two upshots of our discussion of deniability. The first addresses accounts of strategic speech. The second concerns countermeasures against strategic speech.[32]

Lee and Pinker (2010, p. 785) define strategic speakers as seeking 'plausible deniability' and, according to Camp (2018, p. 41), they aim to 'minimize their overall commitments'. All three authors go on to consider different forms indirect speech as means to this end. Our discussion suggests, first, that theories of strategic speech should consider further means to retain deniability. We can retain deniability through direct communication too by e.g. ambiguous wording, a properly chosen environment, or a reputation for overstatement and irony. Second, our account suggests that we widen our conception of the strategic speaker as follows. Strategic speakers seek insulation from rational criticism. Retaining deniability is only one means to this end, retaining untouchability is another. Unlike the means to retain deniability, the means to retain untouchability are often non-linguistic (consider power asymmetries). Sometimes, though, they are linguistic (consider an aggressive discussion style that makes any objection risky), and even if they are not, they are worthy of further consideration. We thus hope that our discussion presents a starting point for future research on strategic speech.

Let us turn to countermeasures against strategic speech. The above considerations suggest that we must distinguish between strategic speakers who seek untouchability and strategic speakers who seek deniability. Limiting untouchability is often a matter of erasing e.g.

---

[30] See e.g. Mendelberg, 2001 on the 'Norm of Racial Equality'.

[31] On certain forms of anti-intellectualism (e.g. Wright, 2011 and McKenna, 2020), knowledge itself depends on social roles and associated conventions. Such revisionary accounts of knowledge would allow one to accommodate sensitivity of deniability to such factors while assuming the previously stated knowledge norms of action and emotion.

[32] We are grateful to an editor and a reviewer for urging us to make these upshots of our theory more explicit.

power asymmetries or fear of conflict, while limiting deniability is a matter of improving one's epistemic position. Our account suggests specific ways to achieve the latter. Consider evidential deniability first. Limiting evidential deniability is primarily a matter of epistemic vigilance. If you pay close attention to contextual cues, your evidential basis may be strong enough to nail down the speaker on a communicated message even if she denies it. Of course, every contextual parameter is potentially relevant for utterance interpretation, and keeping track of everything is impossible. Keen attention to the question under discussion (the 'QUD'; see most prominently Roberts, 2012), however, might go a long way. Suppose, for instance, that someone proclaims 'We should wait for the numbers on excess mortality to come in' in a discussion on deaths related to COVID-19. When the numbers are in, she goes on, 'I never meant to convey that excess mortality would be evidence for COVID related deaths, I only meant that these numbers could tell us more about public health in general'. This denial becomes much harder to maintain if you can point out that the QUD concerned deaths related to COVID. Consider psychological deniability next. We can limit psychological deniability primarily by believing what our evidence tells us to believe, not letting ourselves be fooled into e.g. taking irrelevant error-possibilities seriously. As e.g. Gerken (2017, p. 222) points out, it is psychologically possible to '*critically* assess' situations where an error-possibility has been made salient and thus to realize that the error-possibility in question is either epistemically irrelevant or ruled out by one's evidence. Since critical assessment is cognitively costly, though, it doesn't happen automatically; it must be 'triggered'. It may thus prove effective to adopt a policy to critically assess salient error-possibilities whenever you face strategic speakers.[33] Finally, if we assume that conventions and social roles partially determine the epistemic standard for actionability, we can further limit (evidential and psychological) deniability by changing these conventions and social roles; for instance, by clarifying and adjusting legal or administrative interpretations of the 'reasonable person standard' (see e.g. Lawlor, 2021 for philosophical discussion of what the reasonable person standard could reasonably be). The practical implementation of all of these strategies remains to be explored, but we hope our theoretical framework facilitates this task.[34]

References

Benton, Matthew A. n.d. 'Knowledge norms'. In: *The Internet Encyclopedia of Philosophy*. URL: http://www.iep.utm.edu/kn-norms/.

Berstler, Sam 2019. 'What's the good of language? On the moral distinction between lying and misleading'. In: *Ethics* 130.1, pp. 5–31.

Blome-Tillmann, Michael 2008. 'Conversational implicature and the cancellability test'. In: *Analysis* 68.2, pp. 156–160.

Blome-Tillmann, Michael 2014. *Knowledge and Presuppositions*. Oxford: Oxford University Press.

Boogaart, Ronny, Henrike Jansen, and Maarten van Leeuwen 2020. ''Those are your words, not mine!' Defence strategies for denying speaker commitment'. In: *Argumentation* Latest Articles, pp. 1–27.

Brown, Jessica 2008. 'Subject-sensitive invariantism and the knowledge norm for practical reasoning'. In: *Noûs* 42.2, pp. 167–189.

Buchak, Lara 2014. 'Belief, credence, and norms'. In: *Philosophical Studies* 169.2, pp. 285–311.

Camp, Elisabeth 2018. 'Insinuation, common ground, and the conversational record'. In: *New Work on Speech Acts*. Ed. by Daniel Fogal, Daniel W. Harris, and Matt Moss. Oxford: Oxford University Press, pp. 44–66.

Davies, Alex 2019. 'Testimony, recovery and plausible deniability: A response to Peet'. In: *Episteme* 16.1, pp. 18–38.

DeRose, Keith 2009. *The Case for Contextualism*. Vol. 1. Knowledge, Skepticism, and Context. Oxford: Oxford University Press.

Dinges, Alexander 2018a. 'Anti-intellectualism, egocentrism and bank case intuitions'. In: *Philosophical Studies* 175.11, pp. 2841–2857.

— 2018b. 'Knowledge and availability'. In: *Philosophical Psychology* 31.4, pp. 554–573.

Dretske, Fred I. 1970. 'Epistemic operators'. In: *Journal of Philosophy* 67.24, pp. 1007–1023.

Fricker, Elizabeth 2012. 'Stating and insinuating'. In: *Aristotelian Society Supplementary Volume* 86.1, pp. 61–94.

García-Carpintero, Manuel 2018. 'Sneaky assertions'. In: *Philosophical Perspectives* 32.1, pp. 188–218.

Gerken, Mikkel 2017. *On Folk Epistemology: How We Think and Talk about Knowledge*. Oxford: Oxford University Press.

Goldberg, Sanford C. 2020. *Conversational Pressure: Normativity in Speech Exchanges*. Oxford: Oxford University Press.

Grice, Herbert Paul 1989. *Studies in the Way of Words*. Cambridge and MA: Harvard University Press.

Grindrod, Jumbly 2020. 'Wrongful ways to raise the epistemic standard'. In: *Episteme* First View, pp. 1–15.

Hawthorne, John 2004. *Knowledge and Lotteries*. Oxford: Oxford University Press.

Hawthorne, John and Jason Stanley 2008. 'Knowledge and action'. In: *Journal of Philosophy* 105.10, pp. 571–590.

Jamieson, Kathleen Hall 1992. *Dirty Politics: Deception, Distraction, and Democracy*. Oxford: Oxford University Press.

Khoo, Justin 2017. 'Code words in political discourse'. In: *Philosophical Topics* 45.2, pp. 33–64.

Lackey, Jennifer 2007. 'Norms of assertion'. In: *Noûs* 41.4, pp. 594–626.

Lasonen-Aarnio, Maria 2010. 'Unreasonable knowledge'. In: *Philosophical Perspectives* 24.1, pp. 1–21.

Lawlor, Krista 2021. 'Knowledge and reasonableness'. In: *Synthese* 199.1-2, pp. 1435–1451.

Lederman, Harvey 2018. 'Uncommon Knowledge'. In: *Mind* 127.508, pp. 1069–1105.

Lee, James J. and Steven Pinker 2010. 'Rationales for indirect speech: the theory of the strategic speaker'. In: *Psychological Review* 117.3, pp. 785–807.

Lewis, David 1996. 'Elusive knowledge'. In: *Australasian Journal of Philosophy* 74.4, pp. 549–567.

Maguire, Barry 2018. 'There are no reasons for affective attitudes'. In: *Mind* 127.507, pp. 779–805.

Mazzarella, Diana et al. 2018. 'Saying, presupposing and implicating: How pragmatics modulates commitment'. In: *Journal of Pragmatics* 133, pp. 15–27.

McKenna, Robin 2020. 'Feminist epistemology and pragmatic encroachment'. In: *Social Epistemology and Relativism*. Ed. by Natalie Ashton et al. London: Routledge.

Mendelberg, Tali 2001. *The Race Card: Campaign Strategy, Implicit Messages, and the Norm of Equality*. Princeton, NJ: Princeton University Press.

Michaelson, Eliot 2021. 'Speaker's reference, semantic reference, sneaky reference'. In: *Mind & Language* Early View, pp. 1–20.

Myers-Schulz, Blake and Eric Schwitzgebel 2013. 'Knowing that P without believing that P'. In: *Noûs* 47.2, pp. 371–384.

Nagel, Jennifer 2010. 'Knowledge ascriptions and the psychological consequences of thinking about error'. In: *The Philosophical Quarterly* 60.239, pp. 286–306.

Peet, Andrew 2015. 'Testimony, pragmatics, and plausible deniability'. In: *Episteme* 12.1, pp. 29–51.

— 2018. 'Testimonial knowledge without knowledge of what is said'. In: *Pacific Philosophical Quarterly* 99.1, pp. 65–81.

Pepp, Jessica 2018. 'Assertion, lying and untruthfully implicating'. In: *The Oxford Handbook of Assertion*. Ed. by Sanford Goldberg. Oxford University Press, pp. 829–850.

Pinker, Steven, Martin A. Nowak, and James J. Lee 2008. 'The logic of indirect speech'. In: *Proceedings of the National Academy of Sciences of the United States of America* 105.3, pp. 833–838.

Reed, Baron 2010. 'A defense of stable invariantism'. In: *Noûs* 44.2, pp. 224–244.

Reins, Louisa M. and Alex Wiegmann 2021. 'Is lying bound to commitment? Empirically investigating deceptive presuppositions, implicatures, and actions'. In: *Cognitive Science: A Multidisciplinary Journal* 45.2, pp. 1–35.

Roberts, Craige 2012. 'Information structure in discourse: Towards an integrated formal theory of pragmatics'. In: *Semantics and Pragmatics* 5.6, pp. 1–69.

Rose, David and Jonathan Schaffer 2013. 'Knowledge entails dispositional belief'. In: *Philosophical Studies* 166.S1, pp. 19–50.

Saul, Jennifer M. 2017. 'Racial figleaves, the shifting boundaries of the permissible, and the rise of Donald Trump'. In: *Philosophical Topics* 45.2, pp. 97–116.

— 2018. 'Dogwhistles, political manipulation, and philosophy of language'. In: *New Work on Speech Acts*. Ed. by Daniel Fogal, Daniel W. Harris, and Matt Moss. Oxford: Oxford University Press, pp. 360–383.

Sharot, Tali 2012. *The Optimism Bias: Why We're Wired to Look on The Bright Side*. London: Constable & Robinson Ltd.

Stanley, Jason 2015. *How Propaganda Works*. Princeton, NJ: Princeton Univ. Press.

van Elswyk, Peter 2020. 'Deceiving without answering'. In: *Philosophical Studies* 177.5, pp. 1157–1173.

Viebahn, Emanuel 2020. 'Lying with presuppositions'. In: *Noûs* 54.3, pp. 731–751.

Walton, Douglas 1996. 'Plausible deniability and evasion of burden of proof'. In: *Argumentation* 10.1, pp. 47–58.

Weiner, Matthew 2005. 'Must we know what we say?' In: *The Philosophical Review* 114.2, pp. 227–251.

— 2006. 'Are all conversational implicatures cancellable?' In: *Analysis* 66.290, pp. 127–130.

Williamson, Timothy 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

— 2005. 'Contextualism, subject-sensitive invariantism and knowledge of knowledge'. In: *The Philosophical Quarterly* 55.219, pp. 213–235.

— forthcoming. 'Justifications, excuses, and sceptical scenarios'. In: *The New Evil Demon*. Ed. by Julien Dutant and Fabian Dorsch. Oxford: Oxford University Press.

Wright, Sarah 2011. 'Knowledge and social roles: A virtue approach'. In: *Episteme* 8.1, pp. 99–111.

Zakkou, Julia 2018. 'The cancelability test for conversational implicatures'. In: *Philosophy Compass* 13.12, pp. 1–14.