Conversational Double Bookkeeping

Andy Egan, Rutgers University

Draft of February 25, 2025

1. Introduction

One of the theoretically important things about a conversational situation is that there's a bunch of stuff that the participants in the conversation collectively take for granted. There's a body of information that's *presupposed*. Conversations are guided by, and (at least largely, at least often) aimed at elaborating on, a collective model of a way for things to be. The role of conversational common ground is to determine or characterize what that model is like.

Assertion, on a popular Stalnakerian model, is fundamentally in the business of updating our collective model by adding more information to it – by introducing new presuppositions into the conversational common ground. As the conversation goes on and more assertions get accepted, more stuff gets taken for granted – more stuff gets presupposed. Our model gets more specific, gets more informationally rich, and rules out more possibilities.

In Stalnaker's way of modeling the common ground (I'll sometimes abbreviate to 'CG') this elaboration on our collective model shows up as the context set – the set of possibilities compatible with the conversational presuppositions – getting smaller. In other models it shows up as the set of propositions that are presupposed getting bigger. The conversational common ground is theoretically central to lots of ways of thinking and theorizing about language and communication, both as something that shapes and constrains speakers' conversational moves (by e.g. resolving context dependence, determining whether a potential utterances' presuppositions are satisfied or not, determining what kinds of contributions will be appropriately informative, etc.), and also as something that is shaped by speakers' conversational moves.

There are a number of theoretically noncommittal ways to describe the attitude parties to a conversation take to the common ground – we can say that they take it for granted, treat it as background, presuppose it, for example. And there are a number of theoretically committal options of how to explain just what taking for granted, etc. amount to. A natural thought is that what's common ground is what's common *knowledge* (what all parties to the conversation know,

know they all know, etc.), but that seems too strong, because of the possibility of false and/or evidentially under-supported presuppositions. A natural next thought is that what's common ground is what's common belief (what all parties believe, believe they all believe, etc.). A view that's probably popular enough to count as the received view, is that common belief won't quite do, because of the possibility of accepting for purposes of the conversation, and taking for granted in the relevant sense, things that one doesn't believe. (Maybe the most popular example of this is the case where your uncle inserts some contentious political commentary into the conversation, which is mostly about something else, at the holiday table, and while you don't agree, you're not willing to derail the conversation by objecting, and so you let it slide. Some other plausible cases: I describe a vice-chancellor as a dean, or a golden lab as a golden retriever, in a context where the difference doesn't matter. Or I get a street name a little bit wrong. Or I slightly mangle a famous quotation, or attribute it to the wrong author, or to the wrong book or paper by the right author.) The standard view about such cases is that what happens to the content of the assertion that's been allowed to pass without objection is that it joins the ranks of the propositions that are commonly *accepted* for purposes of the conversation. And it's common acceptance, and not common belief, that underwrites the common ground: What's common ground in a conversation is what's commonly accepted by the parties to the conversation.[1]

Approximately everybody working in theoretical frameworks that include something conversational common ground-ish in their theoretical apparatus has, since Stalnaker 1978, agreed that these kinds of letting-it-slide cases show that the attitude at the bottom level of presupposition isn't belief, but acceptance. And basically everybody working in such frameworks writes in a way that suggests that they think that, while *officially* we can't say that presupposition is common belief, that fact isn't especially theoretically interesting or important. There are countless paragraphs early in philosophy of language and linguistics papers carefully pointing out that officially it's acceptance on the bottom level of presupposition, and then saying that

[1] There's some variation in whether people think that the attitude that iterates on top of acceptance is belief, or if the iterated attitude is also acceptance. (Is it that P is presupposed when everybody accepts P, everybody believes everybody accepts P, everybody believes everybody believes everybody accepts P, etc., or is it that P is presupposed when everybody accepts P, everybody accepts everybody accepts P, everybody accepts everybody accepts P, etc.? Stalnaker REF does the former, and Stalnaker REF does the latter.) I'm going to write as if the iterated attitude is belief, and as if the difference between these options doesn't matter very much. I'm aware of how uncomfortable that is, given the main thrust of this paper. And I will be extremely surprised if I'm not eventually quite rightly taken to task for it. But (a) it makes presentation easier, and (b) I'm actually pretty eager to provoke somebody into explaining to me exactly how things go wrong when you do it the way I'm doing it almost all the time in the main text.

they're going to write, in the rest of the paper, as if acceptance was just common belief, because it makes presentation easier, and whatever, NBD. I have written a bunch of these. Bob Stalnaker has written some of these. (BRING THE RECEIPTS HERE.) It's very likely, if you've written about this stuff at all, that you have too. (Dan, in "People not Contexts", does something a little more subtle: he sets up the picture in terms of acceptance and then talks about belief in selling the explanatory virtues of the picture.)

So the orthodoxy since Stalnaker 1978, as revealed by the way people write their papers about this stuff, seems to be this: Officially, acceptance is not belief, and it's acceptance, not belief, that's (at least) at the bottom level of presupposition (and perhaps is also what iterates). But we can (almost always) safely ignore that, and talk, write, think, and theorize as if presupposition was common belief.

But that's not right. The fact that presupposition is iterated on top of acceptance, rather than belief, is theoretically very important. And keeping the difference between common acceptance and common belief squarely before our minds while we're theorizing about language and communication is helpful in accounting for a number of phenomena that are otherwise potentially puzzling and theoretically recalcitrant.

The plan for the rest of the paper is to demonstrate the importance and utility of being vigilant about the acceptance/belief distinction. In the next section, I'll discuss Elisabeth Camp's work on insinuation, emphasizing what I take to be a central lesson of it: That we need, in order to adequately characterize what's happening in insinuation, to keep track of both an acceptance-based and a belief-based common ground. (This isn't quite the way Camp describes the framework she advocates for, but I think it's a not-too-distorting alternative mode of presentation of at least one central thread of her picture.) I'll then look at a range of other phenomena that, once we've got both acceptance-based and belief-based common grounds in our theoretical apparatus for representing the states and development of conversations, we're in a better position to account for.

2. (One of) the lessons of Camp on insinuation – the acceptance/belief distinction is important, and you need to keep track of both

Elisabeth Camp's (2018) work on insinuation provides a powerful motivation for being careful about the acceptance/belief distinction. In this section I'll present what I take to be some

of the central lessons to be drawn from this work. Then in subsequent sections I'll apply those lessons to accounts of other phenomena.

To get a feel for the target phenomenon, let's look briefly at some of Camp's cases:

> The motorist who says to the police officer who has pulled them over, "Is there any way we can just take care of this right now?", insinuating (but not explicitly asking) *will you accept a bribe?*
>
> The party to a degenerating romantic relationship who says to their partner who is returning home long after the known ending time of a known-to-be-dull work party, "the party must have been really fun, huh?", insinuating *I know that you are having an affair*.
>
> The real estate agent who says to the prospective homebuyers of a racial, sexual, or religious background that is underrepresented and perhaps unpopular in the neighborhood, "perhaps you would feel more comfortable locating in a more . . . transitional neighborhood, like Ashwood?", insinuating *you are not welcome here*.
>
> (Another version of that case: "one thing to consider is that this is a bit of a… transitional neighborhood", said to the local-majority-matching buyers, looking in a neighborhood that's got higher concentrations of non-majority people in it, insinuating *you wouldn't like it here because of all the undesirables*.)

A crucial feature of these cases, which is distinctive of insinuation: there's a message that's intended to be communicated, but also intended to be deniable. ("What? I just said…") One helpful way that Camp describes the phenomenon of insinuation is that it exploits "the gap between what's in fact mutually obvious to the speaker and hearer, on the one hand, and what both parties are prepared to *acknowledge* as mutually obvious, on the other" (Camp 2018 p48). The utterance has a communicative force that's on the official, above-board, mutually-acknowledged record. And it also has another communicative force that's mutually recognized, but not mutually acknowledged. The first, but not the second, is officially on the table to be explicitly taken up, challenged, etc.

A central feature of insinuation is that there is a bifurcation of communicated messages. In general, what happens with insinuation is that there's an official, above-board conversational move whose occurrence, and whose consequences, are fully acknowledged in the conversational record, and also another conversational move whose occurrence and consequences happen "under the table" in the sense that they are not fully acknowledged in the same way as the above-board message. The insinuated message doesn't play the same role as the openly acknowledged message in the regulation of the downstream development of the conversation: It doesn't, for example, provide an antecedent for propositional anaphora, it's not readily available to be targeted with explicit mechanisms of agreement and denial, and the speaker is not conversationally committed to or accountable for it in the same way as they are for the above-board message.[2]

So one thing that insinuation compels us to do is make theoretical room for two different kinds of statuses for communicative actions and their effects to have – some are above board, and fully acknowledged in the conversation, while others are (in a sense that needs defining) under-the-table, and not-fully-acknowledged.

In an important class of cases of insinuation (what Camp calls "deeply shadowed" insinuation), successful communication involves the satisfaction of the following conditions:

- There's one communicated message that's innocuous, explicitly communicated, officially on the table, and easily available for agreement, denial, challenge, complaint, etc.
- There's another communicated message that's potentially occous, that everybody recognizes, everybody recognizes everybody recognizes, etc.
- But that second content has an importantly different status from the first, and a different role in the subsequent development of the conversation. The second content is communicated, but it's not explicitly communicated, not officially on the table, and

[2] A natural question: What way is that, exactly? That's definitely in need of eplanation, and it's going to be controversial. But for now, we don't need to be clear on just exactly what the asymmetry amounts to, so long as we're convinced that there really is an asymmetry. And I think that much is clear, even if the right way to characterize the asymmetry is not.

it's harder to hold the speaker accountable for it, and harder to explicitly agree with, deny, challenge, complain about, etc.

Here's a way to think about this that I think is a mode of presentation of Camp, though it's not quite the one she presents her picture under: There's a gap between what's accepted - and commonly believed to be accepted - and what's believed - and commonly believed to be believed. In successful insinuation the acceptance-based common ground only includes the above-board, mutually-acknowledged content, while the insinuated content winds up as common *belief*, even though it's not commonly accepted. [3]

A couple points to draw attention to:

First, as Camp emphasizes (and Stalnaker notes), this applies pressure to distinguish what's accepted for purposes of the conversation from what's believed from the opposite direction than the letting-it-slide cases – it gives us things that are commonly believed but not accepted, rather than things that are commonly accepted without being believed. In cases of successful (deep) insinuation, we want the acceptance-based common ground to omit some stuff that the parties to the conversation actually believe, rather than wanting it to include some stuff that they don't believe. We want this because we want a body of information that guides (at least some aspects of) the development of the conversation that doesn't include the insinuated content, and doesn't include a record of the speaker's communication of, or commitment to, the insinuated content.

Second, this highlights the need to keep track of two distinct bodies of information playing different kinds of common-ground-ish roles. We need to keep track both of the acceptance-based common ground that includes the official above-board content but not the insinuated content, and also the belief-based common ground that (in cases of successful insinuation) *does* include the insinuated content.[4]

Some crucial lessons from Camp's discussion of (deeply shadowed) insinuation:

---

[3] I'll be focusing on what Camp calls "deeply shadowed insinuation" (Camp, ppXX), because that's the case that applies the most pressure to (a) very clearly distinguish acceptance and belief, and (b) keep separate track of common-ground-ish versions of both. But officially, these cases, in which it's not just that the communication of the insinuated content is deniable, but also that the insinuated content itself doesn't go in the above-board, acknowledged common ground, are a distinctive special case of insinuation.

[4] Qualification about instances of insinuation where there's significant uncertainty about what's being insinuated, and so you probably can't iterate up very high.

1) It's important, in the explanation of why the insinuation is deniable, and why it's hard to directly challenge and/or hold the speaker accountable for it, that there is an official, conversational-behavior-guiding common-ground-ish body of information that includes the above-board content but doesn't include the insinuated content. (That includes, for example, the request to take care of things right now, but *not* the request to accept a bribe.)
2) It's also important to keep track of something common-ground-ish where what's common ground is genuinely believed - you need that in order to keep track of whether the insinuation was communicatively successful or not.

Crucially, the official, above-board common ground fails to include some facts about the speaker's communicative acts and communicative aims that are in fact mutually obvious to speaker and hearer.[5]

So the belief-based and acceptance-based common ground differ both in that the belief-y CG includes the insinuated content and the acceptance-based CG doesn't, but also in that the belief-y CG incorporates the information that the speaker did a communicative move aiming to communicate insinuated message and the acceptance-based CG doesn't. That's why the speaker preserves official deniability, avoids a certain kind of accountability for the insinuated message, and it's hard to take them to task for it. It's also why the audience avoides a certain sort of on-the-record accountability for accepting the insinuated message.

The resulting picture that I take from Camp (though I don't want to attribute all of this to Camp – I'm not sure she'd sign on for all of it. TALK TO LIZ ABOUT THIS BIT):

We keep the picture where assertion is (at least often) primarily aimed at updating the (official, acceptance-based) common ground.

There are then a number of parasitic communicative effects that can follow on from making assertions in contexts, and some of those update the acceptance-based common ground,

[5] I'm deviating from the official Campian picture here: Camp carves out a separate theoretical widget – the conversational record – for this. I, as a card-carrying Stalnakerian, want to fold that into the context set. There is a lot more to say about this, but I don't think it's super important for present purposes. If Camp winds up winning the fight about this, the lesson will be that we need to keep track of one more thing, not that we don't have to keep track of both of the things that I'll be concerned with.

while others update the belief-based common ground. Some will update both, and some only one.

Something that's distinctive about (deeply shadowed) insinuation is that the speaker has a central communicative aim that's supposed to show up only in an update to the belief-based common ground, and not in the acceptance-based common ground.

Moral: There are some important communicative phenomena such that, in order to model them properly, we need to keep track of two distinct CG-ish bodies of information. One has acceptance on the bottom layer of the iterated attitude, capturing what's taken for granted *for purposes of the conversation*, and what's openly acknowledged to be taken for granted for purposes of the conversation. The other has *belief* on the bottom layer of the iterated attitude, and captures what's taken for granted as *actually reflective of reality*. These might diverge by there being some things that are commonly believed but not commonly accepted (as in the case of deeply shadowed insinuation), or by there being some things that are commonly accepted but not commonly believed (as in the case of letting assertions slide).

Once we have this squarely in view and are attuned to it properly rather than sweeping the acceptance/belief distinction under the rug in the first few paragraphs of our papers (again, I am not just casting stones at other people's houses about this - *mea* totally *culpa* here), a bunch of interesting lessons and applications suggest themselves.

3. Nerdy aside in (David) Lewis scholarship

In *Convention*, Lewis has a one-sided story about the conventions constitutive of being a speaker of a language. There is only a speaker-side convention of *truthfulness* - of producing sentences only in certain kinds of circumstances. There is no corresponding audience-side convention about how to respond. That's because Lewis thought that an audience-side convention would have to be a convention of belief, which you could only have if voluntarism about belief were true, and it's not.[6]

[6] In "Languages and Language", Lewis adds an audience-side convention of belief (Lewis calls it a convention of *trust*). He does that because between the writing of *Convention* and of "Languages and Language", he was convinced (he attributes the convincing to Jonathan Bennett) that conventions didn't have to be conventions of voluntary action, so there could be a convention of belief even if voluntarism isn't true. (See Lewis REF pp138-140 for the details.)

The problem here is an artifact of thinking about audience-side uptake only in terms of belief. There's no parallel problem about an audience-side convention about acceptance, because voluntarism about acceptance *is* true. (That's one of the lessons, or maybe one of the presuppositions, of the letting-it-slide cases – we can *decide* to let our grumpy uncle's assertion slide and accept its content for purposes of the conversation.) The problem that led Lewis to initially give a one-sided account of the conventions of language never arises if you think the audience's side of the (system of) convention, participation in which is constitutive of being a speaker of a particular language, is a convention of acceptance.

(Note included not because it's super relevant to the present agenda, but in order to see if I can provoke readers of the paper at its current stage of development to say stuff that I'll learn something from: I think the "everything is about belief all the time" feature of Lewis's thinking about language in *Convention* and "L&L" actually runs pretty deep, and distorts his picture in important ways. One of the probably four or five places where I'd actually be willing to accept a significant-stakes bet on Lewis being wrong about something.)

4. Coping with factual and linguistic errors

4.1 Malapropisms

A framework that's careful about distinguishing acceptance from belief is helpful in theorizing about how we cope with malapropisms.

Sometimes, malapropisms are a serious barrier to communication. If I mix up "extension" and "intension" when I'm talking about semantics or the metaphysics of properties or externalism about belief, especially if I'm talking to an audience that's not already deeply familiar with the relevant ideas, I'm liable to be misunderstood. In general, if my malapropism results in production of a sentence whose literal meaning is also a plausible (by the audience's lights) thing for me to mean, it's going to be hard for audiences to recognize the malapropism and correct for it. And if I commit too many malapropisms in a single utterance, I'm liable to make my utterance uninterpretable.

But famously, our interpretive capacities in the face of malapropisms are quite robust. (See for example Davidson, "A Nice Derangement of Epitaphs".) Very often, we are able to work out exactly, or near enough to exactly for the purposes at hand, what the speaker meant.

Davidson and his examples obviously loom large here, but as I'm writing in New Jersey rather than California, I'll illustrate with some examples from *The Sopranos* instead:

- Mob boss Tony Soprano says to his therapist, discussing his reaction to another character's death, reports, "I was f******g prostate with grief".
- Tony's lieutenant Christopher, discussing a plan to disrupt the operations of a rival organization, suggests that they "create a little dysentery among the ranks".
- Fellow mobster Bobby Baccalieri contributes to a discussion of current events, "You know, Quasimodo predicted all of this".
- Mob boss Carmine Lupertazzi Sr., after revealing that he knows that Tony is seeing a psychiatrist, reassures him, "So what? There's no stigmata these days. My kid saw a shrink…" .
- Boss Carmine Lupertazzi Jr., discussing with Tony a conflict between Tony and another mobster, mentions "this alteration you had with him…".[7]

In all of these cases, the speaker's intended meaning is easy to recover. Let's call a malapropism where the audience has correctly worked out the speaker's intended meaning a *corrected malapropism*. In the case of a corrected malapropism, the audience faces, after they've worked out the speaker's intended meaning, a further question about how to go on in the conversation.

One thing we can do in response to a malapropism like this is to explicitly call it out. But that's conversationally disruptive, potentially hurtful and relationship-damaging, and sometimes (as in some of the *Sopranos* cases) unsafe. Often, it's better to not make a fuss.

In those cases, we can correct for the speaker's linguistic error and allow the conversation to proceed smoothly, as if they had actually expressed what they were aiming to express. Keeping separate track of acceptance and belief provides a nice model of how we do this: We accept, but don't believe, that the word means what the speaker thinks it means. We accept, but don't believe, the meaning postulate that the speaker believes (and also accepts). And so it becomes acceptance-CG that the word has the malapropistic meaning. The resulting context will be one in that's defective with respect to the belief-CG – the speaker thinks that their meaning

[7] Carmine goes on to tell Tony that he is "at the precipice of an enormous crossroads".

postulate is belief-CG and the hearer knows it's not – but in perfect order with respect to the acceptance-CG. And this non-defectiveness in the acceptance-CG is enough to keep the conversation running smoothly.[8]

4.2 Meaning-shifting, neologisms, and uncareful speech

The existence of this capacity for audiences to cope with speakers' unintentional misuse of words, and speakers knowledge of it, gives rise to the possibility of deliberate (and potentially transparent) exploitation of it by speakers. The same capacity that allows us to cope with malapropisms also allows us to cope with, and to innovate with, neologisms. Speakers recognize that hearers have the capacity to work out what an expression would have to mean in order for the sentence in which it occurs to have a meaning that the speaker plausibly intends. Speakers can, and do, exploit their knowledge of that capacity to use existing expressions with novel meanings, and to introduce meanings for novel expressions.

The process is, plausibly, the same. The hearer recovers the speaker's intended meaning, and accepts for purposes of the conversation that the expression has that meaning. The speaker designs their utterance relying on the hearer to do this, and designs it in order to make it easy for the hearer to recover the intended novel or non-standard meaning. Often (but not always) the speaker will intend for all of this to be recognized by the audience. Sometimes (but not always) acceptance of the newly introduced meaning will carry over between conversations, and sometimes (but not always) the usage will spread, and eventually it will be appropriate to believe, and not merely accept, the relevant meaning postulate, because the expression will have genuinely acquired the new meaning.

This seems likely to be the mechanism at work in the origins of many of the neologisms that emerged from online culture – "noob", "to google", "flamewar", the social media use of "viral", "to sealion", etc. A similar mechanism is at work in our ability to understand the title of

[8] Note that there's room for variation in the manner and degree of integration of the hearer's acceptance of the speaker's meaning postulate into their downstream behavior in the conversation. They might merely interpret the speaker's utterances as if the word had the malapropistic meaning, or they might adopt the alternative meaning themselves, and use the word that way in their own speech. (This second thing is probably more likely in conversations in which there's a substantial power or status asymmetry between speaker and hearer, where the speaker is the one with greater power and/or status.) If the hearer isn't willing to adopt the alternative meaning themselves, they are likely to circumlocute as needed to avoid using the misused word, in order to avoid bringing the conflicting, correct meaning postulate to the surface.

Karen Bennett's paper, "Why the exclusion problem seems intractable, and how, just maybe, to tract it". (There are many uses of "ept" and "ert" in the same spirit, exploiting the same mechanism.) I relied on the same interpretive capacity in my use of "occuous" in the section on insinuation. Some of these cases – "tract", "ept", and "ert", for example, are maybe not so neo-. The fake meanings that speakers (and writers) who use them are relying on their audiences to work out are not especially novel. Still, as in cases of more novel kinds of non-standard usage, the speaker is relying on the audience to accept a meaning postulate that they likely do not believe.

Genuine neologism-coining and this kind of familiar as-if etymology shade in another phenomenon of uncareful speech. This ability of hearers to accept alternative meaning postulates in the interpretation of our utterances enables speakers to be sloppy with their language. Often (you see this a lot in ordinary conversation in the wild, and also in a lot of undergraduate writing) people will not be terribly careful about their word choice, using a word that's sort of roughly in the neighborhood of what they mean (you can probably fill in examples from the last time you graded undergraduate papers), and relying on their audience to fill in an appropriate word meaning. This works fine in most ordinary communication, though it doesn't work so well in philosophy, where it happens quite frequently that small differences in meaning are important, which is maybe why we and our students so often get crosswise of each other about making sure that the words they write down actually mean exactly what they want to say.)

4.3 Correcting for non-linguistic error

We also have the capacity to correct for speakers' non-linguistic factual errors in very much the same way. A couple of examples to illustrate the phenomenon:

- I describe somebody sitting at the bar as "the woman drinking the martini", and go on to say a bunch of stuff about which papers she's written, what she thinks about the norm of assertion, and so on. You know that she's actually drinking water out of a martini glass. (Donnellan, Stalnaker)
- I call a particular mountain "Mount Rainier", and go on to describe the locations of a bunch of other things in our environment in terms of their relations to Mount Rainier. You know it's actually Mount Adams.

In both cases, it's likely to be completely clear which person or mountain I have in mind. And in both cases, you have some options about how to respond: One way that you can respond is by correcting me. But my mistake (let's suppose) isn't relevant to what we care about, and it might be time consuming, distracting, or interpersonally awkward to correct me. So another thing you could do is accept (without believing) that the woman in question is drinking a martini, or that the mountain in question is Mount Rainier.[9] Accepting that falsehood will allow a lot of useful, true information to get incorporated into the common ground – about which papers the woman we're accepting is drinking a martini wrote, and what she thinks about various topics, and about where a bunch of other things are relative to the mountain we're accepting is Mount Rainier (and consequently, where those things are relative to each other). It is often a useful strategy to accept, without believing, that the speaker's describing things correctly. By doing so you can frequently allow the conversational common ground to smoothly take on board a lot of (true, useful) information about the mis-described object, and about its relations to other things, without having to derail the conversation to correct an irrelevant error. (See Stalnaker, *Context*, ChN ppXX)

These look like a species of letting-things-slide case – there's something the speaker's asserted (or implicated, or presupposed) that we aren't willing to believe on their say-so, but we let it slide and accept it for purposes of the conversation. These cases are interestingly different from the standard examples of letting-it-slide cases because in these cases the acceptance-without-belief isn't just motivated by avoiding conflict and conversational crisis (a la Farkas and Bruce 2009). It's also motivated by facilitating communication about stuff that we *are* prepared to believe on the speaker's say-so. We're not just letting things slide to avoid conflict – we're letting things slide to facilitate the smooth transfer of additional information.

Accepting that that woman is drinking a martini, or that that mountain is Mount Rainier, gives you a way to cope with my factual error, and to take on board the accurate information that I'm communicating to you by means of my inaccurate description, without shutting down the conversation and holding things up to correct me. Also without forcing us to find some other way to describe the thing or coordinate our attention on it, which might be difficult. It might be a big

[9] Or maybe better, accept that the mountain is called "Mount Rainier". I'll continue to not be careful about this in the main text.

project to correct my error, or it might be that the by far most salient thing about the object we want to talk about is a misleading appearance.[10]

4.4 Exploitation: Referential uses of descriptions and harmlessly uncareful reference-fixing

The fact that hearers have, and use, these kinds of mere-acceptance-based capacities to cope with speakers' misdescriptions and misspeakings also makes available to speakers some fairly sophisticated strategies for communication, which exploit these capacities.

One thing you can do as a speaker is describe something in the environment is a way that you're confident will direct attention properly, but that you're not sure is accurate. You can then (often) rely on your audience, even if it turns out that they know that your description is inaccurate, to accommodate by accepting your inconsequential misdescription of the target object and taking on board the conversationally relevant information about the mis-described object. You might even draw attention to an object by way of a description that you're *certain* is inaccurate, and that you're certain your audience will realize is inaccurate. (There are going to be a lot of interesting distinctions between cases here that I'm going to skip over for now.) This is one way to generate the phenomenon of referential uses of definite descriptions famously introduced by Donnellan (REF). (Note: It is definitely not part of my agenda here to argue that it's the only way, or that the right analysis of the referential/attributive distinction is going to be in terms of acceptance and belief, or anything like that.)

More generally: You can often use a description you lack confidence in to draw attention to the object you want to communicate some information about, and trust your audience to accept that the description applies, and thereby take on board the information about the (potentially) misdescribed object whose transmission is your actual communicative aim. They might accept because they don't know you are, or might be, misdescribing the object. Or they might accept because because they know that you are or might be describing, but they (like you) don't care about the accuracy of the description, as long as it directs everybody's attention to the right object, and so they are willing to accept that the description applies without believing it.

[10] It's easy to imagine this happening in Herb Clark REFs) tangram experiments. Maybe "looks like the wicked witch of the North" for a tangram presenting a witches-hat-and-cartoonishly-large-nose appearance. (There's no wicked witch of the North, it's the wicked witch of the West who looks like that.) Or "looks like Daffy Duck" for a duck-bill-and-sailor's-hat looking one. (It's Donald Duck who wears a sailor's hat.)

5. As-if communicative situations

Sometimes we regulate our conversation as if the conversational situation was different from how we all know we all believe it to be. We all think the actual situation is *this* way (it's belief-based CG that the conversation is F and not G) but it's also important for us to keep track of an alternative conversational situation that we act (for some purposes) is if we were in (it's acceptance-based CG that the conversation is G and not F).

Having this feature of the structure of conversational situations clearly in view is helpful in accounting for a number of interesting phenomena.

5.1 Talking to the Dog

One case of this that will be familiar to many pet owners is talking to the dog (or cat, or turtle, or…). My particular family organization prominently features a dog (her name is 'Bacon'), who is a participant, at least according the acceptance-based common ground, in many (probably most) of the conversations that happen in our house. (Informal surveys of pet owners at philosophy conferences and at the dog park suggests that this is, while not quite universal, also very far from idiosyncratic.) Very, very many conversations in my family proceed as if Bacon were a full participant, even though we are all aware that she is not.

There are of course some cases of talking to the dog where we really are just straightforwardly trying to communicate with the animal, and producing utterances that we expect to have a predictable effect on the animal's mental states and behavior, based on an at least semi-realistic assessment of the animal's actual cognitive capacities. Most instances of "sit", "stay", "leave it", "here", and many instances of "good dog" and "time for walkies" are like this. This kind of dog-directed speech doesn't require any very fancy theoretical apparatus to accommodate.

But many cases of talking to the dog aren't like this. There is also a broad range of cases where, when we talk to the dog, we organize our conversational behavior as if the dog were the kind of agent who could plausibly extract and update with the literal content of our assertions (and perhaps with various implicatures), and could be a genuine participant in the conversation – a participant whose presuppositions help to determine the conversational common ground. These are the cases where we don't just tell her that she's a good dog, but (for example) make complicated comparative claims about her goodness relative to other dogs in the neighborhood,

or relative to certain acquaintances or colleagues, or about how we're going to use her as an example in a paper, and about the details of the philosophical thesis we're going to use her to illustrate. One thing that's distinctive and in need of explanation about these cases is that we are liable to use sentences that the animal is very unlikely, given a realistic assessment of her cognitive capacities, to be able to understand. But our treatment of the addressed animal as a full-fledged participant in the conversation runs deeper than vocabulary choice. We are also, for example, likely to refrain from anaphoric reference to people who we haven't previously mentioned to her, to fill in required presuppositions before saying things to her that depends on them, and so on

To get a useful picture of what's happening in this kind of case of talking to the dog, it's helpful to have the acceptance/belief distinction ready to hand – all of this behavior makes a lot more sense if we think that the speaker is accepting various propositions about the animal's cognitive capacities and beliefs (and probably also desires, intentions, etc.) that they don't actually believe.[11]

So far we've looked at cases in which a single human member of the family is talking to the dog, acting as if she were a kind of agent that we know she's not. And we've seen that, in thinking about what's happening in these kinds of cases, it's helpful to be able to say that the human party to the interaction accepts a number of propositions that they don't believe, and treats some of those propositions as if they were common ground between themselves and the dog. The human speaker is acting as if they were party to a sort of virtual common ground, allowing their linguistic behavior to be governed (in important respects) by a body of information known not to be common ground (either acceptance-based or belief-based) for any actual group. (See Lewis, "Imagined audiences…")

There are also many cases of conversations that include more than one human member of the family, and in which some of our utterances are addressed to Bacon. This kind of case is more theoretically interesting, because it looks like it requires that there is a body of information

[11] Of course there are cases where people genuinely believe the false stuff about their pets' mental lives. And there are cases in which the pet owners' attitudes toward the false claims about their pets' mental lives maybe don't fit nicely in a simple taxonomy of mental states – they're not fully belief-y, but they're more belief-y than any of the other attitudes that we've got convenient names for (acceptance, imagination, pretending, etc.). I'm going to shuffle this under the rug for now but I actually think it's going to wind up being interesting and important. (Also I am pretty sure some members of my family have this kind of fish-nor-fowl attitude toward the false claims about the dog's psychology.)

that we humans are using to guide our conversational behavior – that really is playing the conversation-guiding role of common ground among the humans – which diverges in important respects from what we commonly believe.

The communicative role of these utterances is potentially puzzling. Sometimes, the utterances addressed to the dog are bank-shot communicative moves aimed at other human members of the family. ("Bacon, why hasn't anybody fed you yet?", "Bacon, are you distracting the kids from doing their homework?", "OMG Bacon, why are there so many dirty dishes on the countertop?") In these cases a central communicative aim is that some human participant in the conversation draw some actually-believed conclusion from the fact that the speaker is making as if to say something to the dog.

But it's not all like that. (For example, a lot of the elaborate good-dog stuff isn't like that.) Some of the dog-directed utterances don't come attached with any very obvious human-directed conversational agenda. They seem instead to be part of a sort of collective fiction-building project. (And plausibly a lot of them have subtler kinds of human-directed communicative agendas.)

5.2 Talking on behalf of the dog

Another, deeper level of making as if the conversational situation was other than we actually believe it to be: Talking on behalf of the dog. In these cases, some human party to the conversation will speak for the dog (possibly using a distinctive voice). Some recent examples from my house: "Heyandy, why haven't you taken me for a walk yet?", "Adeline, did you know that I am a good dog?", "Why do humans get lunch but dogs don't?" Again, some of this is bank-shot communication directed at human parties to the conversation. But, at least in our family, not most of it. A lot of it doesn't serve any very obvious human-to-human communicative purpose. It seems to be largely a project of fleshing out a kind of story, and developing a partially fictional character. (Again, informal surveys suggest that we are not outliers here.)

In this kind of case, we regulate our conversational behavior by reference to a body of information that diverges dramatically from what we commonly believe: One according to which the dog is on record with various assertions, proposals about the QUD stack, requests, demands, rejections and acceptances of others' conversational moves, etc. And a bunch of context-sensitive expressions that feature in utterances in the conversation are going to have their context-

dependence resolved by reference to as-if facts about the dog's preferences, beliefs, communicative intentions, etc.

All of this is much easier to make sense of if our theoretical toolkit includes a distinction between what's commonly accepted for purposes of governing the conversation, and what's commonly believed to be true.

Though those are probably my favorite cases of regulating the conversation according to a picture of the conversational situation that's different from the way we actually believe it is, they're far from the only ones, and I suppose it's possible that they're not the most philosophically significant ones. So let's look at some more cases.

5.3 Retraction and deep retraction

Sometimes, in the face of conflict, challenge, or just later reconsideration, we want to take back some previous speech act. Our linguistic practice seems to provide us with a tool for doing this – a speech act of *retraction*. Here is a popular kind of story about the distinctive effect of retraction: It targets a past speech act, and unwinds its characteristic effect. So perhaps retracting a previously accepted assertion of P strikes P from the conversation's presuppositions and unwinds the speaker's commitment to P, a retraction of a question takes the question off the QUD stack, retracting a command takes the commanded action off the addressee's to-do list, etc. (A BUNCH OF REFS) A standard part of this kind of story about retraction is that retracting an assertion doesn't unwind *all* of its effects. If my assertion surprised you, my retraction doesn't unwind your surprise. And in particular, retracting an assertion can take its content out of the common ground, but it can't take the fact that I made the assertion out of the common ground. (More REFs – Kukla, Vermaire, Caponetto, Schlöder et al…)

This kind of speech act of retraction is easier to account for with (probably doesn't work without) an acceptance/belief distinction, because it looks like it depends on having a body of information with a common-ground-ish status that we can update voluntaristically in response to retracton, based on whether we accept/reject the retraction.

Another phenomenon that the acceptance/belief distinction is helpful in accounting for another, deeper kind of retraction that also seems to be possible. (This phenomenon is on my radar due to a very helpful question from a graduate student at MIT who I frustratingly haven't been able to identify since in order to credit them properly.) Sometimes a speaker will say

something, come to regret it, and say something like, “please forget I said that” or “please pretend I didn’t say that”. This is a bid for *deep retraction* – for excision of not just the speech act’s conventional effects, but the fact that the speech act occurred, from the conversational record. This is, somewhat surprisingly, a sort of request that it seems as if we can sometimes go along with.

Here is a natural account of what happens when we do: we shift to an acceptance-based common ground that doesn’t include the fact that the deeply retracted speech act occurred. When you say something that you regret, you can’t really ask people to *believe* that you didn’t say it. Belief doesn’t work that way. But you can ask them to *accept* that you didn’t say it – you can ask them to excise the fact that you said it from the conversational record, to strike it from the model of the conversational situation that guides how the conversation moves on from there. When somebody attempts a deep retraction and everybody goes along, the acceptance-based CG gets updated so that according to it, the retracted utterance never occurred. (Even though, in the usual case, everyone still remembers that the speech act *did* occur, thinks everybody else remembers it, etc. – so its occurrence is still common belief.)

5.4 An alternative to forward-looking models of speech act typing

Another place where the idea of merely-accepted models of the conversational situation can be constructively applied is to the cluster of phenomena that have motivated a number of theorists to go in for forward-looking and outward-looking models of speech-act individuation. (Some prominent developers and advocates of such views include Rae Langton, Quill Kukla, Laura Caponetto, Bianca Cepollaro, and Lucy MacDonald (REFS)).

The central thought behind this family of views is that whether a particular kind of speech act was successfully performed, and/or what kind of speech act was performed when somebody made a particular utterance, can depend on what happens later in the conversation, after the time of the utterance – in particular, on the responses and downstream actions of other parties to the conversation.

The phenomenon that these sorts of views are responding to is that the role that a particular utterance plays in guiding and constraining the downstream development of the conversation, and downstream non-conversational actions of the parties to the conversation,

really does seem to be dependent on the kind of reception that it receives from its audience.[12] The downstream conversation-guiding role of an utterance does seem to be susceptible to being pushed around by later actions by other parties to the conversation. One natural response is to say that, since the conversation-guiding role of a speech act is determined by the type of speech act that it is, speech act typing had better be sensitive to the later actions of other parties to the conversation.

The OG version of this is Rae Langton (REF) on blocking. Langton's proposal is that it's possible for people other than the speaker, acting after the original utterance has happened, to make it the case that a speech act that was intended to be of a particular type T is not, and never was, of type T. Here is a central motivating example for Langton: I try to give an order to a group that you're a member of, acting as if I've got the relevant kind of authority. If nobody complains, I gain the right kind of authority by accommodation, and I've managed to successfully issue an order. But if you challenge me – "who are you to order us around?" – then the accommodation doesn't happen, and I don't gain the kind of authority that I need to have in order to successfully issue an order. And so, I fail to issue an order. Your later action blocks my earlier speech act from succeeding in being the kind of speech act it was setting out to be. On account of your later authority-sapping objection, my speech act is not, and never was, an order.

A related family of examples, and an account of the nature of speech acts with the same forward-looking, response-dependent feature, comes from Quill Kukla (REFs): On Kukla's view, success in launching a speech act of a particular type is dependent on audience uptake. The speaker hasn't successfully produced a speech act of illocutionary type T unless the audience recognizes the speech act as of type T. As Kukla puts it, "details of how a speech act receives uptake help determine what kind of speech act it is". (REF, pp).

One of Kukla's central examples also involves attempts to issue commands: Celia is a female manager in a factory with an almost exclusively male workforce. Her job description says that she has the authority to give orders to the workers in the factory. Celia tries to give them orders - please unload that truck, etc. She gets low compliance from her attempts to give her workers orders about what to do. One reason why this might happen is that, on account of sexist

[12] At least, this is a phenomenon that these accounts are responding to – I don't want to speak for all of the motivations of all of the advocates of views in this family.

assumptions and expectations in place in the workforce, the workers interpret her attempted orders as requests.

Kukla says: When Celia attempts to tell her workers what to do, her speech act only succeeds in *being* an order if it secures *uptake* as an order. So by only doing request-type uptake, the workers make it the case that Celia doesn't successfully give them orders. She only manages to make requests.

Another example of the kind of forward- and outward-looking taxonomy of speech acts that I have in mind is Caponetto and Cepollaro (REF) on *bending*. The phenomenon they're responding to is the fact that, after an initial speech act, performed with a particular intention on the part of the speaker, it's possible for an audience member to issue a distorted behavioral response to the speaker's intended move – responding to the utterance as if it were of a different type than the one the speaker intended – and thereby change the interpretation the original utterance receives, and the role that the utterance plays in the way the conversation goes on from there.[13]

On Caponetto and Cepollaro's view, the reason why the audience's subsequent response changes the interpretation and downstream role of the original utterance is that it changes what kind of speech act the original utterance was (all along) an instance of. A speech act that *would* have been of type T1 is instead, because of the facts about the audience's later actions, of type T2.

The idea here is similar to Langton's account of blocking, in that whether an earlier speech act is of a particular type can depend on later responses from the audience. In Caponetto and Cepollaro's target cases, though, the later action doesn't cause the earlier speech act to misfire – instead it changes the kind of speech act it is.

Here is one of their central examples (REF, pp):

> Anna, Jason, and a few colleagues of theirs are brainstorming about potential invitees for a graduate roundtable. Jason says, "My supervisor would avoid inviting, you know, affirmative action students". Anna perfectly gets that Jason is

[13] There are definitely connections here with Camp (REFs) on the various kinds of pedantry that audiences are likely to go in for in less than fully cooperative conversations – seminar participants, help getting clear on the relations here would be great!

> implicitly suggesting that they do the same, and yet she goes on replying, "I know. It's terrible how racist some professors are around here". After a moment of bewilderment, Jason mumbles a 'yeah'. The organization process goes on with no regard for the bigoted suggestion to exclude students from underrepresented groups.

Their diagnosis: Anna's move reconfigures Jason's original speech act, so that it wasn't, after all, a proposal to exclude members of underrepresented groups. "Acting as if your interlocutor's move were less prejudiced than it actually is can transform it into a different, less prejudiced, contribution" (pp).

So on C&C's account of bending, what kind of speech act was performed depends on how it gets bent later. And so, as in Langton's and Kukla's accounts, what the type of a speech act is might not be knowable on the basis of, because it might not be metaphysically determined by, just the facts up to and including the time when it happens.[14]

Lucy McDonald (REF) offers a "collaboration theory" of illocutionary force, which I'll conclude with because it's an example of an especially aggressively forward-looking story about speech act typing:

> an utterance has illocutionary force φ
> iff both (a) and (b) occur:
> (a) The hearer communicates to the speaker that she interprets the utterance as having force φ
> (b) The speaker then communicates to the hearer that she accepts the hearer's interpretation of the utterance as having force φ
> (MacDonald REF, pp)

[14] C&C build the normative valence into their analysis of bending – it's only bending if it's an ameliorative response to an attempt to launch an objectionable speech act. I'd prefer, for my theoretical purposes, to carve up the territory differently, using 'bending' to talk about the general phenomenon of after-the-fact pushing around of speech act types, and then talking about positive and negative instances of bending. And in the main text I won't draw attention to the normative-loadedness of 'bending' as C&C use it. But this is ultimately a terminological fight that's not especially interesting or important.

This is a picture of what it takes to be an assertion that P, or a command to perform action A, or whatever, that makes speech act typing dependent on later conversational events in an even starker way than the other views we've looked at so far.

Before moving on to offer an alternative way to understand the conversational phenomena that motivate these kinds of forward- and outward- looking accounts of speech-act individuation, I want to emphasize a couple of points:

First, there clearly is an interesting, important, and potentially theoretically puzzling phenomenon (or cluster of phenomena) that they're responding to – the way conversations actually work clearly does provide a market for this kind of view. The role that a particular utterance plays in governing how the conversation develops really does depend, in the ways these kinds of views identify, on what happens later, including on audience members later responses. And making membership in the speech act types that determine/explain an utterance's role in the development of the conversation depend on temporally downstream conversational actions and events really does promise to explain why this is so. My and Celia's attempted orders don't play an order-like role in how later events play out because they weren't actually orders. And what made it the case that they weren't orders was the blocking and/or distorted uptake that happened after the attempted order. Jason's attempt to propose eliminating minority candidates from consideration doesn't play a proposal-to-eliminate-minority-candidates role in the development of the conversation because, on account of Anna's subsequent successful bending move, it wasn't actually a proposal to eliminate minority candidates.

I also want to emphasize that there's nothing conceptually or metaphysically incoherent or problematic about these proposals. There are ever so many types, membership in which is forward- and outward- looking. Being popular is unproblematically outward-looking. And, if we accept that there are some facts about the future, there are lots of unproblematic forward-looking categories. Being long-lived is unproblematically forward-looking. And there are, assuming that there are facts about the future, straightforward metaphysically unproblematic facts about which of the people of our acquaintance are the long-lived ones. It's just that it's often hard to tell who those people are, because membership in the category is determined by facts about how things go for them in the future that we have, at present, poor epistemic access to. In the same way, and more relevantly for present purposes: We make a lot of decisions every day. Some of them are trivial, and some of them are consequential. But it's hard to tell which is which at the time when

we're making them, because whether a decision is consequential or not depends on what happens later. But there's not, in virtue of the fact that whether a decision is consequential or not is hostage to what later events, any big mystery or incoherence in the notion of a consequential decision. There are tons of forward-looking categories for things – in particular, for actions, conversational or otherwise – to fall into, and their forward-lookingness isn't any kind of metaphysical or conceptual strike against them.

So if there's something wrong or theoretically undesirable about these kinds of views, it's not that the taxonomy of speech acts they propose is metaphysically or conceptually objectionable. The taxonomy is in good order, and there are facts about which utterances fall into which of the categories that feature in the taxonomy. For any given utterance, there's definitely (assuming there are facts about the future) a fact about whether it's a command-as-Langton-understands-it, or a proposal-to-discriminate-as-Caponetto-and-Cepollaro-understand-it. The kind of considerations that ought to lead us to resist going in for such a view (if we wind up deciding that we should resist) will instead be considerations of theoretical utility – is the forward-and-outward-looking taxonomy of speech acts the one that's most helpful in thinking about the roles of particular utterances in the dynamics of conversation, or is something else more helpful?

I think that an account that makes use of a taxonomy of speech acts that is not forward-looking, together with a picture of context that clearly differentiates between an acceptance-based and a belief-based common ground, is better suited to explain the phenomena that motivate forward-looking accounts.

On this sort of account:

A speech act happens at t1, and is of whatever type it is. Call it type F. That type is fixed entirely by facts about events up to the time of the speech act. (Maybe entirely by facts about the speaker – for example, about their communicative intentions – up to t1. But I don't want to commit to this – I want to focus on the forward-lookingness rather than the outward-lookingness for present purposes.)

In the standard case, all parties to the conversation both believe and accept that the speech act is of type F (and adopt the relevant iterated attitudes about both). So it becomes acceptance-CG, as well as belief-CG, that the speech act is of type F. Everyone then proceeds to run downstream conversation on that basis.

Here is something that can then happen later: Somebody acts in a way that would be appropriate if the speech act were of some other type G. Maybe everybody else goes along. Now we reconfigure our acceptance state - the model of the conversation we regulate conversational behavior on the basis of - so that according to it, the speech act was always of type G. And we go on, running the downstream conversation on that basis.[15]

First point to emphasize: We don't (anyway, we needn't) update our belief-states in the same way. We will (or at least we're likely to) wind up in a situation where the acceptance-based CG says that my earlier speech act was G, the belief-based CG says that it was F, and the different aspects of our downstream behavior proceed accordingly.

Second point to emphasize: We don't need a forward-looking account of speech act individuation to do this. When we update our acceptance state to one according to which the speech act was F rather than G, what we're doing is changing which backward-looking type we're accepting that the speech act was an instance of. (So for example, after the change, we're likely to wind up accepting (but probably not believing) that the speaker's intentions were of the kind that would make their speech act an instance of F.)

This kind of account doesn't require that the downstream blocking, bending, uptake, or negotiation actually changes what kind of speech act occurred. Instead, it changes what we collectively accept (for purposes of the conversation) about what kind of speech act occurred. But it still lets us accommodate the phenomenon where the blocking, bending, uptake, or negotiation changes the role that the speech act plays in guiding the post-blocking/bending/etc development of the conversation.

It also lets us retain the ability to explain the role of the speech act in what happened in the conversation *before* the blocking etc. occurred. For a while, conversation was governed by an acceptance state that matched our beliefs about (and the truth about) the kind of speech act that was performed. Then some moves got made that changed how the speech act was represented in our acceptance-based model. And that changed the role the speech act played in guiding conversational behavior after the bending etc. happened.

6. Assertions not aimed at belief

[15] This is set up as a direct alternative to bending – one would have to change some details for alternatives to the other kinds of forward-looking phenomena.

6.1 Jokes and Stories

People frequently tell jokes and stories. Often, a lot of the things people say in the course of telling a joke or a story aren't true. Often, a lot of the things people say in the course of telling jokes and stories are known, by both speaker and audience, not to be true. Often, neither speaker nor audience cares if they're true. And that's all fine. It's all fine because jokes and stories don't (typically) aim at truth. (Whatever "it's funny because it's true" means, it's not that.)

One response to the fact that people say all kinds of false, and known to be false, stuff while telling jokes and stories, and nobody seems to mind – nobody objects, or gets confused, or accuses the speaker of norm-violations – is to say that they're not *really* saying false things, because there is, for example, a ubiquitous unpronounced 'fictionally' operator at work in these kinds of speech. Another is to sharply differentiate the kind of speech act we're performing in joke- and story- telling from the kind we're performing in paradigmatic cases of inquiry-focused, truth-seeking conversation where we're really trying to zero in on the actual world by eliminating alternative possibilities. (REFs)

Another (I think better) option is to just differentiate acceptance from belief. Then we can say that what people are doing in the course of joke- or story- telling is off-the-shelf assertion, and what assertion does is (as usual) update an acceptance state. Then we draw attention to the distinction between acceptance and belief, and categorize joke- and story- telling contexts as contexts in which it's understood between speakers and hearers that what's accepted is, for the most part, not to be believed. (As opposed to paradigmatic Stalnakerian finding-the-actual-world conversations, that allow (at least mostly) for free exportation from acceptance to belief.)[16] The assertions you make in the course of telling a joke or a story are updating an acceptance state, in a context where there's usually little or no temptation to export from acceptance to belief. (Probably we often leave it an open question whether and how much of we're accepting is to be believed - this seems like a nice thing to say about the "don't care if it's true" rather than "take for granted that it's false" cases.)

6.2 Teaching

[16] See Yalcin (REF) on conversational tone, and Harris and Unnsteinsson (REF) on conversational genre. More on this in section 7.

A similar kind of situation arises very frequently when teaching undergraduates. Very often, I'll say stuff to my students that I don't think is true, but that I think provides a good model for them to have in their heads for some purpose. Maybe (for example) it's close enough to true, or it generates a lot of good predictions without being super complicated, or it illustrates an important idea more clearly than the truth, or maybe it's a view that it's helpful for students to inhabit for a while before being pushed out of it by counterexamples.

That stuff definitely doesn't become belief-based CG, because I don't believe it. I also (often) don't especially want them to believe it. In a lot of these kinds of cases, I'd ideally want them to accept it, and to believe that it stands in some useful modeling relation to the truth. But often it's also fine if they do believe it - it's a good enough model for their purposes, and we can fix it later if and when they're in grad school.

I don't think this is just me. It seems certain that science teachers do this all the time. They'll teach the simplified not quite true version, and then fix it later. Sometimes teachers will explicitly flag that they're doing this, but they don't always, and they don't have to.

In a lot of cases, what you want as a teacher in those cases is acceptance, plus some belief that what they've accepted stands in some interesting relation to the truth (you don't want them to treat it as just a story). Often it's fine if they accept it and believe it, and also fine if they accept it only as an approximation or a useful model.

6.3 Discussion in active research fields

Something similar happens (very plausibly) in active research fields.

In active research areas where things are unsettled and research is very much still in progress, it frequently happens that the evidence just is not in a state to warrant anybody having very high confidence in any specific hypothesis. There will, often, be no hypothesis such that a rational assessor of the evidence could muster a credence in it high enough to clear any plausible threshold for belief. (It's pretty easy to convince oneself that this is the situation in a lot of areas of philosophy, for example. But many areas of experimental science are like this, at least in some stages of their development, as well.)

But still, people working in these areas go around advocating for particular hypotheses and giving papers making claims about stuff. And that's good – one of the ways progress happens in these fields is by people adopting research programs in which they advocate for some

hypotheses rather than others, even in the absence of the kind of evidence that would rationally support belief in the hypotheses they're advocating for. (Also, for distribution-of-scientific-labor reasons, it's good to have a bunch of people working on hypotheses other than the one that's best supported given the current state of the evidence (REFs – Kitcher, Strevens, etc.))

One response to this, developed in a series of papers by Will Fleisher (REFs), is that the attitude that underwrites (a lot of) assertion, publication, etc in many active research areas isn't belief. Fleisher calls the relevant attitude *endorsement*. The details of endorsement's similarities to and differences from belief aren't relevant for our purposes. What's important for our purposes is that, if some account of this type is right, and assertion etc. in these kinds of fields is supported by, and aimed at producing, not belief but the distinct attitude of endorsement, this gives us another family of cases of assertions not aimed at belief.

If I'm giving a talk in the kind of active research field Fliesher is concerned with, what's happening is that I'm asserting a bunch of theoretical claims that I merely endorse. Even if you wound up believing them, they wouldn't enter into the belief-based common ground, because I don't believe them. But also, my aim is almost certainly not to get my audience to believe my theoretical claims, but at most to get them accepted for purposes of the conversation, and for my audience to export them from acceptance, not to belief, but to endorsement.

I think this Fleisherian model of discourse in active research fields, or something very like it, is probably correct. But even if it's not, it's certainly possible, and a discourse of this type would be in good order. And in order to properly model what's happening in this kind of conversation, it's extremely helpful to have at one's disposal both an acceptance-based and a belief-based common ground.

7. Conversational tone and conversational genre

Seth Yalcin (2007), in a paper that takes the acceptance/belief distinction seriously, introduces the helpful notion of *conversational tone*:

> An attitude is the *conversational tone* of a group of interlocutors just in case it is common knowledge in the group that everyone is to strike this attitude towards the propositions which are common ground. (Yalcin 2007 p1008)

Conversational tone is something that varies across conversations, and its central theoretical role is to fix the attitude participants are expected to take to what's commonly accepted. In the kinds of serious-inquiry-focused conversations philosophers are often thinking of when they're theorizing about language, the conversational tone is *belief*. But, as Yalcin notes, the conversational tone is not always belief: "the conversational tone may be belief, or suspicion, or supposition, or high credence that, or ironic non-belief, etc., depending on the interests and purposes of the interlocutors" (Yalcin 2007, p1008)

This is a very helpful contribution, which promises to help us in getting clear about how to think about the way that participants in a conversation export from acceptance to other attitudes. But while the introduction of the idea of conversational tone is important progress, tone is not a sufficiently versatile tool to model the way in which acceptance interacts with other attitudes. That's because there's typically no uniform attitude we're supposed to take toward everything that's accepted in a particular conversation. (Yalcin actually acknowledges (same paper, same page) that it might not be enough, voicing a suspicion that we'll sometimes need to say that a single conversation has more than one tone.)

A helpful example is a phenomenon that Fleisher (REF) draws attention to about the research talk case: Suppose that Fleisher is right, and that the attitude scientists take toward their own theories is endorsement rather than belief. Now I'm giving my talk about some contentious research area, advocating for hypothesis H. I motivate H by telling my audience about some experimental results E that I got by running some experiments in my lab, and pointing out H's ability to predict and explain E. What should we say, again assuming that Fleisher is right, about the conversational tone of my talk?

I want all of the stuff I asserted – about the truth of H, about the experimental protocols we used in my lab, about getting results E, and about H's explanatory value – to be accepted. I don't want all of it to wind up believed - even I don't believe H. What I'm hoping for from you in the audience with respect to H is, at most, endorsement. But I don't want you to merely endorse the things I said about my experimental results. *That* I want you to believe. And I had better believe them in order to assert them during the talk – while endorsement is enough to support my assertion of H, I had better not assert E unless I believe it. So there's no uniform fact about what attitude we're supposed to take, or expecting each other to take, toward what's accepted. In Fleisherian scientific discourse, some of what's accepted is to be endorsed, and

some of it is to be believed. (In general, the contentious theoretical claims are to be endorsed if they're accepted, while the experimental results that are put forward to support the theoretical claims are to be believed.)

Here is a less theoretically fancy case, right next door to a paradigmatic kind of "letting things slide" case: At a family wedding, I'm driving from the ceremony to the reception with my uncle with whom I am politically at odds, and he is giving sound directions and saying a bunch of crazy bananas stuff about politics. It's all getting through to the conversational common ground, because I'm letting the politics slide. The political stuff is not going to, is certainly not commonly expected to, and is likely not expected by anybody to, export to belief. (My uncle could very easily be fully aware that I'm just letting the political assertions slide.) But the stuff about which turns to take and which landmarks to look out for *is* supposed to export to belief.

The same goes for stories. Mike Birbiglia is a comic storyteller who does one man shows in which he tells highly personal stories that are also very funny. It would be unreasonable for his audience to expect that everything in a Mike Birbiglia story is 100% true. But his audience *does* expect that stories don't pull completely apart from the truth – they expect that at least the central features of the core narrative of the story are true, and that the events the story is based on really do occupy the same kind of emotional role in Birbiglia's life that he characterizes them as having.

So the facts about which, of the things that you accept as a result of what Mike Birbiglia says during a show, you export to belief, and which you strike some other attitude toward, are going to be pretty complicated. Also, those facts will plausibly not be the same for Birbiglia as they are for other people doing storytelling shows - different performers are going to set up different expectations about the relations between their stories and the truth.

In general: The stuff that's in the acceptance-y CG isn't going to be uniform with respect to which other attitudes you're supposed to take toward it if you accept it. So there's not, in general, going to be any such thing as *the* conversational tone. The facts about how we do, and how we expect each other to, export from what's accepted into various other attitudes (including but not restricted to belief), are going to display a much more complicated structure.

Here is what I think is a better option: Replace *tone* with something like conversational *genre*. Shen-yi Liao, in a discussion of the phenomenon of imaginative resistance, notes that different literary genres are associated with different expectations about how what's true in the

fiction – what's *fictional* – is related to what's actually true. These differences can be thought of in terms of the import/export principles that govern which actual truths it's safe to import into the fiction, and which fictional truths it's safe to export back to one's beliefs about the non-fictional world. For example, in realistic fiction it's quite reliably going to be the case that whatever's actually true about physics, or biology, or geography, is true in the fiction as well. (That's one way that lots of things get to be fictional that aren't explicitly stated by the author – for example, that Sherlock Holmes has a heart. It's also how you can learn some science and history from reading the right kinds of fiction.)

In works of science fiction, however, one can't in general import the actual scientific facts into the fiction – it often happens that, for example, the facts about the possibility of faster-than-light travel are different in the fiction than they are in the actual world.

Literary genres are distinguished, in part, by the kinds of connections that obtain between what's true in fictions in that genre and what's actually true. And the different import/export rules that are expected to apply to works in different genres (and sub-genres) differ from each other in quite complicated, subject-matter-specific ways.[17] For example, a time travel story may allow for free import of scientific facts *except* for the ones immediately involved in allowing for the possibility of the mechanism of time travel that's employed, and also of facts about history up until a specific date when the first interference by time travelers occurs. Many works of historical fiction are governed by import/export rules according to which what's true in the fiction aligns with what's actually true as far as large scale, sweep-of-history matters, but diverges with respect to the specifics of the lives of (and maybe the existence of) the central characters in the story.

I think it's helpful to think of the relation between acceptance and belief on this model. In engagement with fiction, there are (following the influential model proposed by Walton (REFs)) a number of things that are to be imagined. Some of the things that are to be imagined are also to be believed, and some of the things that we believe are to be imagined, even though they're never explicitly discussed by the author. But, obviously, not *everything* that's to be imagined is to be believed.

---

[17] I'm following Liao in talking about the kinds of relations between fictionaiity and truth that are *expected* to obtain in different genres, rather than directly about the relations that in fact obtain in different genres, because it's definitely possible for authors to create works that defy the standard expectations for a genre, and also because there is quite a lot of intra-genre variation in the precise details of the import-export principles governing particular works.

The rules or expectations governing the connections between imagination and belief in engagement with fiction are going to be complex. For one thing they'll be highly variable across fictions, in ways that are sensitive to, among other things, literary genre. For another thing, the connections will, for any particular work, by highly variable across different subject matters. There won't be any such thing as *the* attitude (other than imagination) one is supposed to take toward what's imagined when engaging with a fiction. Some is to be believed. Some is to be believed-possible. Some is, probably, merely to be imagined. (These are not the only possibilities.)

Similarly, the rules or expectations governing the connections between acceptance and belief in engagement with a conversation are going to be complex. For one thing they'll be highly variable across conversations. For another thing, the connections are, for any particular conversation, likely to be highly variable across different subject matters. Just as the import/export principles governing the connections between imagination and belief in engagement with fiction are complex and genre-dependent, so too are the connections between acceptance and belief in conversation.

(Note for seminar participants – I was originally inclined to propose using 'conversational genre' just to talk about collections of import/export principles, and to propose replacing Yalcin's 'conversational tone' with 'conversational genre'. And I was disappointed to learn that Harris and Unnsteinsson were also using 'conversational genre', but using it to talk about something else. I now think it's probably more helpful, because more closely parallel to the case of literary genre, to use 'conversational genre' the way that Dan and Elmar do, and talk about import/export principles for acceptance being genre-dependent in the conversational case in very much the same way that import/export principles for imagination are in the literary case.)

8. Allusions to Cans of Worms

8.1 Two is not going to be enough

For sure the two-way distinction here isn't going to be enough. Josh Dever on letting insinuated content slide. Also probably need >1 accepted-for-different-purposes CG-ish bodies of information for conversations where something complicated is happening – e.g. RPGs and improv classes, probably a bunch of Goffman presentation-of-self phenomena and the kinds of

situations Berstler is talking about in “The Structure of Open Secrets”. So it’s all going to be kind of complicated and gross, in its full-dress version.

8.2 Diagonalization goes wonky when it’s not a sure thing that the same CG is going to play the background-for-interpretation role and the target-of-update role.

8.3 What’s this attitude of acceptance, exactly? How does it fit into a plausible picture of psychology? And what differentiates it from belief – in particular, from fragmented belief with an appropriately circumscribed behavior-guiding role?

9. Conclusion

Coming in the next version, probably.