On Successful Communication, Intentions and False Beliefs

by

MATHEUS VALENTE (iD)

Universitat de Barcelona

Abstract: I discuss a criterion for successful communication between a speaker and a hearer put forward by Buchanan according to which there is communicative success only if the hearer entertains, as a result of interpreting the speaker's utterance, a thought that has the same truth conditions as the thought asserted by the speaker and, furthermore, does so in virtue of recognizing the speaker's communicative intentions. I argue, against Buchanan, that the data on which it is based are compatible with a view involving Fregean modes of presentation. In the second part of the article I critically discuss Unnsteinsson's claim that communicative success depends on the absence of contextually salient false distinctness beliefs about the subject matter of the conversation. I argue that this thesis leads to clearly counterintuitive consequences and that no fundamental role must be given to the presence or absence of false distinctness beliefs in one's account of successful communication. The upshot is that we should stick with Buchanan's criterion. I conclude by employing Strawson and Recanati's concepts of linking and merging to show how the criterion I favour is compatible with the fact that, when subjects hold no relevant false distinctness beliefs, communicative success does not seem to be disrupted by the hearer seemingly failing to recognize the speaker's intentions.

Keywords: communicative success, co-reference, communicative intentions, singular reference

1. Introduction

As a first pass, a communicative event is any that essentially involves the transmission of information from a source to a receiver by means of a perceptible signal. This broad conception of communication might be useful for one purpose or another, but it will not lend itself to much philosophical analysis. The first restriction we shall make is to concern ourselves exclusively with linguistic communicative events between a speaker and a hearer, where a speaker asserts some thought by means of a representation (e.g., an utterance, an inscription, etc.) which is then supposed to be interpreted – successfully or not – by the hearer. The philosopher of communication then wonders about what the conditions on the hearer's interpretative process are for it to have been a successful one, that is, for that event to have been one of successful communication. Even at that narrower level of abstraction, one cannot help but realize that linguistic communicative events of that type can be evaluated according to distinct and independent standards of communicative success which do not

always output the same results. Take, for example, an utterance of the following sentence:

(1)   That man [pointing to a photograph of Nietzsche] is the greatest philosopher who ever lived!

There are at least three different ways in which an interpretation of an utterance of (1) could fail. First, a hearer could fail to assign the same standing meanings to the utterance as the speaker does. So, for example, if the confused hearer thinks "philosopher" means *musician*, while the speaker knows that it just means *philosopher*, the first will have failed to understand what the second meant and, thus, their communicative event would not have been successful. Second, a hearer could fail to grasp *what was said* by the utterance even after correctly assigning its words with the right standing meanings. That is most obviously the case when context-sensitive expressions are present. Thus, the hearer could understand that "that man" is a complex demonstrative expression usually accompanied by some demonstrative act, but mistakenly believe that the speaker is pointing to a person standing across the room from them, and not to the photograph. Failure to understand the speaker's demonstrative intentions would also compromise the communicative event's success, regardless of the correct assignment of standing meaning (character). Finally, the hearer could get everything right at the level of standing meaning and what was said, but still fail to draw an important implicature intended by the speaker. That would be the case, for example, if the hearer were oblivious to the fact that (1) was meant as an ironic utterance, possibly in a context where Nietzsche is having his philosophical reputation harshly scrutinized.

Not only do we seem to have at least three different notions of communicative success corresponding, respectively, to the correct grasp of standing meaning, what is said and conversational implicatures, but the success conditions for these levels seem to be thoroughly independent of each other. Pagin (forthcoming) suggests the following communicative exchange as an instance of success with respect to correct grasp of conversational implicature, but failure with respect to standing meaning and what is said:

(2)   Anna: Would you like to go to the movies?

Bob: I have to go to her [pointing at Claire] bank.

Anna would be able to draw the implicature that Bob is too busy to go to the movies regardless of whether she thinks Bob is talking about a financial institution or the riverside and regardless of whether she thinks he is pointing at Claire or the person standing right beside her. In other words, (2) would be an instance of communicative success with respect to the implicature that Bob is busy even

in a case where Anna fails to grasp what is said by Bob or even the meaning of one of the words of the utterance.

That all being said, if we intend to investigate communicative events as regards the criteria for their success, the first thing to do is to restrict our focus. In this article, we will only be concerned with success with regard to standing meaning and what is said.[1] Additionally, our focus will be almost exclusively on communicative events employing paradigmatic types of referential expressions: proper names, demonstratives and indexicals. Some of what I will say is supposed to apply for linguistic communication in general, but making sure that all of our examples contain similar types of expressions will help us get a clearer sense of the notion of communication that is at stake. Finally, we will limit ourselves to assertoric exchanges, i.e., those involving properly assertoric uses of sentences in the declarative mood.

Thus, from this point onward, "communication" will be used to refer to the more specific type of interaction between a speaker and a hearer with respect to the meaning of and/or what is said by an utterance containing a referential expression produced by the former. Analogously, by "communicative success" I will mean success with respect to grasping what was said by the speaker and/or assigning the appropriate meanings to the relevant utterance. That notion of communication (and of communicative success) might not *perfectly* map onto all ordinary-language uses of the words "communication" and "understanding", but I intend it to approximate one of their central and most important ones. Indeed, the main objective of this article can be described as an attempt to shed light on and trim the edges of our folk-notions of communication and understanding.

2. Communicative Success and Identity of Truth-Conditions

Criteria for communicative success should minimally spell out necessary conditions for a speaker to have successfully communicated a thought to a hearer by means of an utterance. It is a platitude that one of the most general aims of communication is enabling a reliable transmission of information between thinkers; thus, it is plausible that an instance of successful communication should at the very minimum guarantee that the belief asserted by the speaker and that subsequently entertained by the hearer necessarily have the same truth-value. In other words, the simplest criterion of successful communication is one according to which success is determined by the hearer acquiring a thought with the same

---

1   These would seem to boil down to the same thing for *prima facie* non-indexical expressions such as names.

truth-conditions as the one the speaker intends to convey; let us call it the *identity of truth-conditions criterion (C1):*

(C1) There is successful communication between speaker S and hearer H iff H entertains, as a result of interpreting S's utterance, a thought that has the same truth-conditions as the thought asserted by S's utterance.[2]

   C1 is as desirable as it is simple. If adequate, it would allow us to account for communicative success without having to resort to any theoretically loaded notions. This very simple view is, unfortunately, subject to devastating counter-examples. One can easily conceive of cases where a hearer ends up entertaining a thought with the same truth-conditions as that asserted by the speaker but where, intuitively, communication has not been successful. Let us call these cases "Loar-cases", since their inspiration comes from Loar (1976). In Loar-cases, it seems that the failure of communication has got to do with the fact that the sameness of truth-conditions between the speaker and hearer's thoughts is not generated by the usual process of correct interpretation.

2.1 Modes of presentation and communicative intentions: Buchanan against Loar

Here is Loar's (1976) original story:

(Loar-case 1) Suppose that Smith and Jones are unaware that the man being interviewed on television is someone they see on the train every morning and about whom, in that latter role, they have just been talking. Smith says, "That man is a stockbroker", intending to refer to the man on television; Jones takes Smith to be referring to the man on the train. Now Jones, as it happens, has correctly identified Smith's referent, since the man on television is the man on the train; but he has failed to understand Smith's utterance.

   In Loar-case 1, Jones interprets Smith's demonstrative use of "he" as if it were anaphoric on previous uses in the conversation, thus leading us to the intuitive conclusion that they fail to communicate successfully regardless of the sameness of truth-conditions between their thoughts. Loar notoriously took these cases to present a refutation of direct reference views of singular thought, according to which the thought asserted by an affirmative utterance of "*a is F*" is exhausted by the singular proposition that *a is F.* Against that view, Loar argued that, on top of the singular proposition asserted by a speaker in a communicative event, we need to take into account *the modes of presentation* (henceforth MOPs) by means of

---

2   Let's ignore the issue of necessary propositions for simplicity's sake. However, if one really wants to take care of those, then the principle could be restated as "there is successful communication between speaker S, who produces an utterance ascribing property F to object a, and hearer H iff H entertains, as a result of interpreting S's utterance, a thought that ascribes property F to object a".

which that speaker conceives that proposition. By introducing MOPs as components of what is meant by an utterance over and above what they refer to, Loar naturally reached the view that successful communication requires not only sameness of truth-conditions between speaker and hearer's thoughts, but sameness of MOP as well.

As Buchanan (2014, p. 57) remarks, Loar's argument in favour of a MOP-based view of thought and communication depends on the unspoken assumption that understanding an utterance "is simply a matter of recognizing *what* the speaker asserted". If one assumes that and agrees that Jones recognized exactly the same singular proposition semantically associated with Smith's utterance, then it naturally follows that Smith's utterance must have semantically expressed some other content.

It does not take much to see that this line of argumentation does not stand on its own feet. Indeed, it is strikingly similar to the influential Fregean-inspired (and by now generally agreed to be faulty) argument in favour of Senses according to which differences in cognitive value (e.g., co-referential identity statements with different informative potential) must be accounted for by the postulation of fine-grained semantic values that determine the reference of a representation. The point is not that those arguments are invalid; the problem is the presumption that their conclusion is inevitable. Indeed, a significant part of the philosophy of language literature of the last century can be taken as going back and forth over the point that there are alternative explanations to the Fregean *data* which stick to the view that the content of a singular utterance is exhausted by a singular proposition.[3] The main difference between the classical Fregean argument and Loar's is that the latter does not seem to depend on the informativity of identity statements but merely draws upon our intuitions about communicative success – still, they are analogous in fixing the insufficiency of referential content by means of the postulation of a finer type of content.

In the present article, I am interested in the notion of communicative success in its own right, and not only in the prospect of using it to motivate a general account of thought. That being said, it is of utmost importance to look for the simplest and most neutral way of cashing out the lessons of Loar-case 1, and Buchanan's view might very well be a good place to start.

Buchanan's main point is that we can account for what goes wrong in Loarcase 1 by appealing to an independently motivated account of the role of communicative intentions in conversations – one that is so plausible that everybody is more or less obligated to accept it anyway – and thus completely bypass the idea that the content of thoughts and assertions goes beyond what they refer to. The

---

3 Salmon (1986) is a good place to begin surveying the literature on Frege's puzzle.

idea, in a nutshell, is that "the kind of misunderstanding that Loar has called to our attention shows that there is *some aspect* of the speaker's communicative intentions that her hearer is failing to recognize" (Buchanan, 2014, p. 64). In other words, Buchanan is suggesting that the failure of Smith and Jones's communication can be accounted for by the fact that, even though Jones reached a thought with the same truth-conditions as Smith's, he did not do so by means of properly recognizing Smith's communicative intention to refer to the man on the television. On that view, successful communication requires that the hearer token the same singular proposition as the speaker *as a result of recognizing the proper inferential process intended by the speaker*:

(C2) If there is successful communication between speaker S and hearer H, then H entertains, as a result of interpreting S's utterance, a thought that has the same truth-conditions as the thought expressed by S's utterance and, additionally, H does so in virtue of having recognized S's relevant communicative intentions.[4]

Buchanan's argument is as simple as it is convincing. Surely any theorist recognizes the role that speaker-intentions and their recognition by hearers have in the proper functioning of communication. Even the most die-hard Fregean philosopher (who believes that interpretations are a matter of assigning MOPs to symbols) has to tell a story about how people come up with a particular MOP during an interpretative process, and that story will most likely involve the recognition of the communicative intentions of the source to be interpreted. Thus – and this is where I am in complete agreement with Buchanan – a view that accounts for communicative success by means of the recognition of the speaker's communicative intentions makes the postulation of MOPs redundant.

Now, even though I agree with Buchanan's argument against Loar's conclusions, I think its impact should not be overplayed. It is important not to forget that the target of Buchanan's criticisms is a very specific type of Fregean account of communication, that is, an account which takes MOPs to be constitutive of what is meant by singular utterances. On the other hand, many recent philosophers defending views which they consider to be Fregean-inspired hold a much more deflationary attitude towards MOPs and instead assume that a MOP can be anything as long as it is able to play a set of interrelated semantic/epistemic roles

---

4   This is a simplified version of the principle that Buchanan (2014, pp. 63–64) in fact goes on to defend. His principle includes the concept of an *ib-feature*, i.e., a feature of the speaker's utterance which she intends that the hearer use as a basis for her interpretation. This detail is not relevant for the concerns of the present article. It should also be noted that C2 only states necessary, but not sufficient, conditions for successful communication. This is in line with how Buchanan himself frames his own discussion, although Bach and Harnish (1979, ch. 5) can be said to have defended the same principle in its biconditional form. The weaker version is good enough for our present purposes: it is the necessity of the principle, not the sufficiency, that will be put into question in the next section.

even if, at the end of the day, they do not end up being the kinds of things which are expressed by our assertions nor the building-blocks of our thoughts.[5] Thus, if one is not careful, the difference between a MOP-based view, deflationarily construed, and the type of view which Buchanan wants to defend could boil down to a few distinct terminological choices. To evade these complications, let us continue to use "MOPs" with its more restricted meaning, i.e., as contents of thoughts and assertions that outstrip what they refer to.

That being said, Buchanan does indeed present a strong case against MOP-based accounts of successful communication. His argument, however, is not a direct one but one based on the greater simplicity of another available account. It is thus important to keep in mind that a MOP-based explanation of communicative success is still a coherent option on the table, even if it is not the most parsimonious one around. Buchanan, however, also intended to oppose that weaker compatibility claim – and that is where our disagreement lies.

2.2 Misinterpreting a drawing: against Buchanan

On top of arguing that MOPs are not needed to account for Loar-cases, Buchanan (2014, p. 62) claims that the MOP-based view of successful communication cannot account for certain cases which are both (i) analogous to Loar-case 1 in the relevant respects and (ii) suitably explainable by his preferred account. If correct, this could very well constitute a direct argument against any criteria of successful communication based on the postulation of fine-grained semantic contents. I do, however, believe that Buchanan's argument fails and can be adequately answered. This is the case the author comes up with:

(Buchanan-case) In observance of a religious holiday, Smith is forbidden to read, write, or speak for the day. Because Smith is looking so bored, his friend, Jones, tells Smith he will take him to a movie, but they need to decide what to see. It is mutual knowledge between them that a cowboy movie entitled *Flat-top Mountain* is one of the many movies playing at their local Cineplex. Smith grabs his notebook and draws a mountain (in clear view of Jones), intending to communicate thereby that he would like to go to see *Flat-top Mountain.* Jones, however, mistakes the drawing for one of a cowboy hat, and infers thereby that Smith would like to go to see *Flat-top Mountain.*

---

5   One recent example is García-Carpintero's (2016) Fregean-inspired presuppositional view, according to which MOPs are like presuppositions in that they are conveyed (presupposed), but not directly expressed, by means of our assertions and thoughts. In his most recent work, Recanati (2016, pp. 145–146) also adopts a kind of presuppositional account of MOPs according to which the role that modes of presentation are supposed to play are executed by the vehicles of thought – which he calls "mental files" – and presuppositional content they convey depending on the type of epistemically rewarding relations on which they are based.

According to Buchanan (2014, p. 63), "what Smith intended to communicate, and all he intended to communicate, was *that he wants to go to see Flat-top Mountain*". The author then goes on to claim that, unlike in Loar-case 1, here we seem to have no clear candidate for a MOP that the speaker had in mind and that the hearer failed to grasp, that is, "it is completely unclear what the MOP-involving proposition could be in this case" (Buchanan, 2014, p. 63) – but if that were true, then the MOP-based view would seem to have no resources for explaining why their communication appears to have been unsuccessful. The intentions-based view, on the other hand, easily explains what is going on in the story by invoking Smith's intention that Jones recognizes his drawing as a draw-ing of a mountain (and not of a hat).

As a first attempt at a reply, one could complain that Buchanan fails to consider the most obvious Fregean response: Jones thinks of the movie *Flat-top Mountain* via the MOP *the salient movie related to cowboy hats* instead of the MOP that Smith had in mind, *the salient movie related to a flat mountain.* That response could be based on the fact that pictures, just as much as linguistic expressions, can be associated with multiple semantic values, including reference and modes of presentation.[6] Thus, the response would continue, communication fails because Smith's drawing refers to *Flat-top Mountain* via the mountain-MOP but is taken by Jones to refer to that movie via the distinct cowboy hat-MOP. Jones would have gotten the right referent by means of the wrong MOP.

To be sure, this is not the only account of Buchanan's case available to a friend of MOPs. Indeed, even if, for whatever reason, one had suspicions about the idea of Smith's drawing being used to singularly refer to *Flat-top Mountain*,[7] a MOP-based theorist could suggest the following alternative account: Smith's drawing is a genre picture, a kind of depiction whose content is general (like a picture of a horse, but of no particular horse), and its content is the property of being a moun-tain.[8] It is by means of figuring out that his drawing is a drawing of (the property of being) a mountain that Smith hopes Jones will infer which movie he wants to

---

6   Hyman (2012) defends the application of the sense/reference distinction to pictures and other forms of depiction.

7   One could, for example, rephrase Buchanan's story so that the intended communicated content were completely general, e.g., Smith could have intended to express his adoration for Western movies by means of drawing a mountain (commonly associated with the landscape of movies set in the Wild West) whereas Jones thought that he was doing so by means of a drawing of a cowboy hat (equally associated with that type of movie). In that variation of the story, there would seem to be no singular content for the drawing to be a representation of.

8   Analogously, one could say it represents the concept of a mountain. Genre pictures are unlike other paradigmatic instances of depiction, such as portraits, in that they refer to kinds of objects but not to par-ticular instances of them. This means that a simple account of pictorial representation according to which they represent what they resemble would not be immediately applicable to them. For more about theories of depiction and genre pictures, see Hyman (2012).

see. In Fregean terms, this means that the MOP of Smith's drawing is that of the property of being a mountain. One can, if one wants, say either that this MOP refers to the property of being a mountain or that it does not refer at all. Not much will hinge on this, since Jones's mistake can be accounted for by the fact that he fails to correctly recognize the drawing's MOP and, instead, takes its MOP as being that of the property of being a cowboy hat. If one believes that the correct MOP either referred to a property or that it did not refer at all, then one gets the additional verdict that Jones not only failed to assign the correct MOP, but also the correct referent of the drawing. In any case, there seems to be no difficulty for a MOP-based view to account for why Jones has failed to understand Smith.

In summary: Jones misinterprets Smith's drawing as if it were of (the property of being) a cowboy hat, instead of (the property of being) a mountain. But misinterpretation just is, at least in a Fregean framework, the incorrect assignment of MOP to a representation. Thus, their communication fails because Jones fails to associate the correct MOP to Smith's drawing. Furthermore, given some additional assumptions about the drawing's MOP's referent, not only does Jones assign the wrong MOP to Smith's drawing, but also the wrong referent. The case can then be said to be analogous to one where a subject interprets an utterance of "the [river]bank is muddy" as saying that some financial institution is dirty, and, *contra* Buchanan, not analogous to Loar-case 1 (where one at least gets the right referent).

That, I take it, is a natural description of the Buchanan-case which makes it clear that it is compatible with a MOP-based account of communication. Buchanan might have been assuming that MOPs somehow apply only to singular expressions, but that does not correspond to how the notion is employed in practice. Indeed, it seems that for any case that involves some type of misinterpretation – be it of an utterance, a drawing, or any other type of representational act – there will always be some easy way to account for it by means of a MOP-based view.[9]

2.3 Interim conclusion: successful communication requires intention recognition

This is what we have so far: C1 fails as a criterion of successful communication because of cases such as Loar-case 1. In response to that, Loar suggested a MOP-based account of communication which, although coherent and able to account for all the relevant *data*, was shown by Buchanan to be excessively committal. Against Loar, Buchanan suggests C2, a principle which seems to be able

---

9   It seems that the same cannot be said for cases – such as those presented by Byrne and Thau (1996) in discussion with Heck (1995) – where there is no misinterpretation but where the correct interpretation is achieved by sheer luck or coincidence. It is less clear whether the MOP-based view possesses enough resources to deal with these cases or whether it would need to be supplemented with some additional epistemic constraints. I thank an anonymous referee for this commentary.

to do all the work that Loar's MOP-based view was supposed to do and for a much lower theoretical price.

This could have been the end of the story if it were not for a recent paper by Unnsteinsson (2018a) from which one can extract the following criticism: Buchanan's appeal to intention recognition is not adequately explanatory since it does nothing to explain why our intuitions on communicative success seem to change so drastically when we let go of one of the assumptions made in the previous Loar-cases – namely, that hearer and speaker are ignorant of some relevant identity fact. According to Unnsteinsson's proposal, one's account of successful communication must give a much more central role to the presence or absence of false distinctness beliefs such as those that Smith and Jones hold about the man on the TV and the man on the train.

3. The Relevance of False Distinctness Beliefs: Against Unnsteinsson (2018a)

Notice what happens to one's intuitions about communicative success in Loar-cases as soon as we assume that the relevant subjects are enlightened with respect to the salient identity fact in the story:

(Loar-case 2) Suppose that it is mutual knowledge between Smith and Jones that the man being interviewed on television is someone they see on the train every morning.[10] Smith and Jones have not been engaging in any conversation when Smith abruptly says, "That man is a stockbroker", while pointing to the man on television; coincidentally, at that very moment Jones happened to be remembering the last encounter they had with that man on the train. Influenced by his own memory and failing to notice that Smith was pointing to the television, Jones takes Smith's use of "That man" as intending to bring about a memory of that man on the train. Now Jones, as it happens, has correctly identified Smith's referent, since the man on television is the man on the train – but did not do so by means of the proper method.

Do we want to say that Loar-case 2 is a case of successful communication? Unlike Loar-case 1, where our intuitions weighed heavily in favour of

---

10   In an earlier version of this article, it was merely assumed that Smith and Jones knew (and knew that each other knew) that the man being interviewed on television is a man they see every morning. However, as an anonymous referee aptly pointed out, this formulation of the story would be susceptible to familiar problems in case one of the subjects falsely believed that the other falsely believed that he falsely believed that the man on the television was distinct from the man on the train. The concept of *mutual knowledge*, taken from Schiffer (1972, p. 30), was specially devised to take care of similar types of cases. A and B mutually know that p if and only if (i) A and B know that p, (ii) A knows that B knows that p, (iii) B knows that A knows that p, (iv) A knows that B knows that A knows that p, (v) B knows that A knows that B knows that p, (vi) A knows that B knows that A knows that B knows that p, and so on *ad infinitum*.

communicative failure, Loar-case 2 is constructed in such a way that our intuitions, by themselves, do not seem to point *decisively* either way. As Unnsteinsson (2018a) suggests, a good way to begin answering this question is by reflecting on what Smith's reaction would be if he noticed that Jones never realized he was pointing at the television. Here is one natural possibility about what would happen: nothing. Smith would be completely indifferent since his most general objective, e.g., letting Jones know some guy is a stockbroker, would be fulfilled regardless of the inferential process by means of which Jones reached that thought. To be fair, he could feel a bit annoyed and make it clear to Jones that he was actually pointing to the man on the television, but that would certainly be a tad pedantic, and a proper response from Jones would be "So what? It's the same person!"

Unnsteinsson takes these observations as evidence that the failure of communication in Loar-case 1 has got more to do with the fact that speaker and hearer hold a false belief about the distinctness of the object referred to by the speaker in the conversation (that the man on the train is distinct from the man on the television) than with the way by means of which the hearer interprets the speaker.[11] The author's suggestion is that "it is in the nature of the speech act of singular reference that having specific false beliefs about identity can make it impossible for a speaker to perform the act properly", such that "lacking such false beliefs at the time of the utterance can be a condition on the proper functioning of the underlying mechanism of singular communication" (Unnsteinsson, 2018a, p. 5).

One natural way to read Unnsteinsson's proposal is that lacking false distinctness beliefs must be taken as a necessary condition for communicative success.[12] A view of referential communication as a speech act with underlying normative constraints which preclude its proper functioning by confused speakers could have promising features. Nonetheless, I think it leads to clearly counterintuitive consequences.[13] For starters, one could insist that Loar-case 2 is an instance of

---

11 For the sake of simplicity, I will just assume that the subjects in our thought experiments do in fact hold the relevant false distinctness belief, but all of what I am going to say (as well as all of what Unnsteinsson says) is compatible with them merely suspending belief on the issue.

12 Unnsteinsson is more closely concerned with providing necessary conditions for the success of the speech act of singular reference and does not discuss criteria for successful communication *per se*. My reading of that author is thus committed to a certain extension of his view under the plausible assumption that, if false distinctness beliefs were to be disruptive of the success of the speech act of singular reference, then they would also be disruptive of the possibility of understanding them.

13 Three things should be noted about my discussion of Unnsteinsson's view. First, I am merely concerned with criticizing one particular thesis – that lacking false distinctness beliefs is necessary for communication – that is congenial, but, without further argument, not necessarily essential, to what this author defends in a brief discussion note. As is often the case with these short notes, Unnsteinsson might not have had enough space to develop his views fully and, in any case, it is not clear whether his other points are not compatible with my criticisms. Second, there is much of interest in his overall project –

communication failure, regardless of Smith's indifference. One could defend that view by claiming that sometimes our extra-communicative objectives are fulfilled even when, strictly speaking, our interlocutors have not grasped what we said. One could reinforce that view by arguing, as Evans (1982) and Heck (1995) once did, that the *raison d'être* of communication is the transmission of knowledge, something which is bound to be absent in every case where an element of luck is involved in the hearer's interpretation.

It is doubtful whether this knowledge-based view of communication would end up sounding convincing to everybody – as Pagin (2008, p. 30) critically says, "the claim that transfer of knowledge is 'the purpose of communication' strikes me as a piece of metaphysical speculation". In any case, there are even more pressing reasons why that criterion fails: it predicts communication failure every time speaker and hearer have a false identity belief, even when that belief is playing no immediate role in the conversation. It strikes me as obviously true that we can sometimes successfully refer and communicate about people about whom we hold some irrelevant false distinctness belief: Lois Lane successfully talked about Superman almost every day even though she would laugh at the idea that his true identity was that of her boyfriend, Clark Kent.

In a couple of places, Unnsteinsson (2018a, pp. 3, 5) suggests that the problem of false distinctness beliefs arises only when they are contextually salient or relevant. That sounds like a plausible enough idea because it is hard to deny that there are cases where subjects successfully communicate regardless of holding false distinctness beliefs that are contextually irrelevant, e.g., ancient Babylonians were able to communicate successfully about Venus in various contexts (during the day by means of "Phosphorus", during the night by means of "Hesperus") even though none of them knew that Hesperus is Phosphorus. In other words, that ancient Babylonians falsely believed that Hesperus and Phosphorus were distinct stars obviously did not interfere with the fact that they often had conversations about Venus.

What about Smith and Jones's belief that the man on the train is not the man on the television in Loar-case 1? One could argue that this was a contextually relevant belief since they were talking about the man on the train right before Smith

---

that of providing a speaker-based intentionalist theory of reference – that remains untouched by my discussion, such as his discussion of the optimal conditions of the speech act of singular reference and of the viability of an intentions-based view of reference in the face of the conflicting intentions that confused speakers have. I refer the reader to Unnsteinsson's other works (2016, 2018b) for more details. Finally, it might be that Unnsteinsson has a different methodological objective than I do. While I am concerned with an account of the folk notion of communication, he might be concerned with constructing a technical notion able to play a set of explanatory roles in his more general theory of reference. If that is the case, it is possible that he would try to resist arguments based on folk intuitions (see especially Unnsteinsson, 2018b).

made an utterance referring to the man on the television. So far, so good: Unnsteinsson's necessary condition is disrespected and communication does indeed fail. However, re-run Loar-case 1 with a slight modification: assume that Jones notices that Smith is pointing to the man on the television and thus that Jones does not make any interpretative mistake this time. I find it obvious that their ensuing communication would be successful regardless of their false and contextually salient distinctness belief – still, Unnsteinsson's criterion would output the opposite prediction. In summary, subjects can communicate successfully in the presence of false and contextually salient distinctness beliefs.

Maybe I am not being completely fair to Unnsteinsson's notion of contextual relevance. Perhaps what Unnsteinsson means is that false distinctness beliefs only disrupt referential communication when the speaker intends to refer to one of the flanks of the belief and the hearer takes her to be referring to the other. In other words, perhaps Unnsteinsson is thinking of a case where (i) there are two ways W^1 and W^2 of singularly referring to the same individual, (ii) speaker and hearer falsely believe that W^1 and W^2 are ways of singularly referring to distinct individuals, (iii) the speaker intends to refer to that individual by means of W^1 and (iv) the hearer takes the speaker to have referred to that individual by means of W^2. For illustration, W^1 could be *demonstratively referring to an individual as the man on the television* and W^2 *demonstratively referring to an individual as the man on the train*.[14] Plausibly, if all four of these conditions are satisfied, speaker and hearer will have failed to communicate successfully.

Even if it rings true, this idea will not be of much help to the view that lacking false beliefs is necessary for successful communication, since then the resulting view would simply boil down to the claim that false distinctness beliefs disrupt communication when the hearer fails to recognize the speaker's communicative intentions. That is just what is happening in the situation previously described: the hearer thought that the speaker intended to talk about the man on the train while actually she intended to talk about the man on the television. But that is already accounted for by means of C2. Why would we additionally need to give a central role to false distinctness beliefs in our criterion for successful communication if what really matters is whether the hearer gets the speaker's intentions?

In summary, Unnsteinsson's idea was supposed to help us understand why our intuitions become uncertain in cases where enlightened subjects misinterpret one another, such as in Loar-case 2. Against this author, I have argued that it is implausible to claim that subjects cannot successfully communicate about x when they hold false distinctness beliefs about x (even when these are contextually

---

14 Another example would be: W^1 is referring to Venus by means of "Hesperus"; W^2, by means of "Phosphorus".

relevant), i.e., lacking false distinctness beliefs is not necessary for communication success. Furthermore, I have not found any adequate way of fleshing out what Unnsteinsson means by "contextually salient" without making his criterion redundant.

Up until now, one could complain, I have only shown that lacking false distinctness beliefs is not necessary for communicative success. But that is compatible with a modification of Unnsteinsson's view that still gives a central role to the presence or absence of these beliefs. According to this hypothetical view, while it is not necessary that speaker and hearer be enlightened for them to communicate successfully, if they are indeed enlightened, then it is sufficient for the success of their communication that the hearer entertains a thought with the same truth-conditions as the speaker. In other words, according to this view, successful communication is particularly easy for enlightened subjects, i.e., C1 is true for pairs of speakers and hearers that hold no false distinctness beliefs about the matter at hand.

I will finish this section by arguing that C1 is false even for subjects that hold no false distinctness beliefs. This should be seen as providing more justification for accepting my claim that false distinctness beliefs should not have a central place in an account of successful communication. A quick example is enough to suggest why that is the case. Imagine, again, a case where it is common knowledge between Smith and Jones that the man on the train is the man on the television, but where Jones interprets Smith's utterance of:

(3) That man [pointing to the man on the television] is that man [intending to refer to the man they saw on the train]

as if Smith had said:

(4) That man [pointing to the man on the television] is that man [still pointing to the same man on the television]

I take it for granted that this would not be a case of successful communication. What this shows is that, even if speaker and hearer are enlightened, successful communication will not come so easily: C1 fails even for that restricted set of subjects. In conclusion, I have argued *contra* Unnsteinsson that one should not hold the absence of false distinctness beliefs to be a necessary condition for successful communication. I have then made the additional point that, even if one is only concerned with subjects who hold no false distinctness beliefs about the subject matter of their conversation, it is not the case that entertaining a thought with the same truth-conditions as the speaker's is a sufficient condition for communicative success. My overall aim was to argue for the thesis that false distinctness

beliefs should not play a central role in one's account of successful communication.

Regardless of all that, we still have not explained one of the main intuitions behind Unnsteinsson's suggestion, that is, we still need to provide some kind of explanation for why our intuitions change so clearly as soon as we re-describe Loar-case 1 as Loar-case 2, i.e., when ignorance gives way to enlightenment. If my objective of rescuing C2 is to be achieved, there is still some work to be done.

4. Enlightenment, Linking and Merging

As I have remarked, one could accept all of the previous considerations but still wonder why it is that, when we enlighten the hearer and speaker in a Loar-case, as we did in Loar-case 2, our intuitions suddenly become much more sympathetic towards the verdict that they have successfully communicated regardless of the hearer failing to grasp the speaker's communicative intentions properly. That would mean C2 is subject to clear counter-examples.

Loar-case 2 might be less than ideal as a case study because it involves an explicit element of luck in the way the hearer reaches his interpretation. Regardless of what one thinks about communication and luck, it will be useful to analyse a variation of the same case which does not include an accidental stroke of luck.[15] It will then be harder to deny that we can have communicative success even when a hearer seems to have failed to grasp the speaker's communicative intentions:

> *(Loar-case 3)* Suppose that it is mutual knowledge between Smith and Jones that the man being interviewed on television is someone they see on the train every morning and about whom, in that latter role, they have just been talking. Indeed, they have been talking about that man for hours, constantly alternating between referring to him by pointing to his image on the television or by remembering their encounters on the train. Smith says, "That man is a stockbroker", while pointing to the man on television; unaware of Smith's pointing gesture, Jones takes Smith to have been invoking a memory of one of their encounters on the train (perhaps of a day on which the man was dressed as a stockbroker). Now Jones, as it happens, has correctly identified Smith's referent, since the man on television is the man on the train – and was not merely lucky to have done so.

---

15   I have been purposefully avoiding any discussion of the role of luck in Loar-cases and its relevance for criteria for successful communication. As far as the recent literature on the topic goes, there just is no consensus on the question about whether successful communication can be lucky. One tradition that includes Evans (1982), Heck (1995) and Peet (2017) says it cannot. Another tradition including Byrne and Thau (1996), Paul (1999) and Pagin (2008) says it can. At the end of the day, I do not think it is necessary, for the objectives of this article, to try to settle on an answer to this question here.

It is hard to deny that Jones has understood what Smith said and that their communication was successful. Whether it matters or not, Jones was not merely lucky to identify the referent of Smith's utterance. On the contrary, it is obvious that he knew who Smith was talking about, regardless of failing to notice his pointing at the television. Regardless of all that, one could argue that C2 would *prima facie* entail that their communication was not successful since the hearer seems to fail to grasp the speaker's communicative intention. In the final section of this article, I will provide a brief explanation of how one could account for Loar-case 3 without departing from C2.

There is an easy way to explain the communicative success in scenarios such as Loar-case 3 where the subjects are enlightened but still *seem to* misinterpret each other. In order to do so, I suggest we borrow the notions of linking and merging from Strawson (1974, pp. 51–56) and their further developments by Recanati (2012, pp. 42–53). The idea, in a nutshell, is that identity judgements be understood as enabling the flow of information between representations that the subjects previously thought referred to distinct people. For the purposes of this article, linking and/or merging representations can be conceived as processes that have an effect on a subject's communicative intentions with these representations. When a speaker acquires an identity belief about what she previously thought were two distinct things, the conditions of fulfilment of her communicative intentions about that thing will also expand so as to cohere with her new state of mind. Let us see how that would work in practice.

Assume that there is a time *t* previous to the interaction described in Loar-case 3 when Smith and Jones are still unsure about whether the man on the train is the man on the television. At *t*, Smith and Jones take the world to be such that there are two distinct individuals satisfying those predicates and strive to isolate the information they acquire about each of them. At that point in the story, when Smith makes an utterance about, e.g., the man on the train, the fulfilment of his communicative intention depends on Jones taking him to be remembering the man on the train (and not, for example, pointing at the man on the television). When these subjects learn that the man on the train is the man on the television, they are rationally required to update their mental states by means of the mechanisms of linking and merging. In other words, this means they no longer have a reason for keeping the information about the man on the train insulated from information about the man on the television. In response to that, they should either create a bridge between their information repositories or merge them, so that all of the information can be stored in one and the same place.[16] From that

---

16   I do not intend my folder, repository or compartment-talk to be taken as anything more than meta-phors for how we organize the information we possess.

point onward, Smith's communicative intentions become fulfillable by his inter-locutors taking him to be intending to refer to the man on the train. Thus, his communicative intentions expand so as to encompass distinct ways of referring to what he now knows is one and the same person.

Now we can see why false distinctness beliefs matter to our intuitions of com-municative success. When we acquire information about one and the same thing but are unaware of that fact, we are rationally required to keep that information in distinct "compartments" until we make sure that they do, indeed, concern the same thing. When we discover that we are receiving information about the same thing from different non-coinciding means, we then have the option of merging the information into one and the same compartment or linking the information contained in distinct compartments in order to allow for their free flow (by expanding the fulfilment conditions of our communicative intentions). To be fair, Unnsteinsson could try to incorporate these linking and merging mechanisms into his account – but if he did, then he would owe us an explanation for why false distinctness beliefs are fundamentally, and not only derivatively, explanatory.

Thus, when Smith and Jones become aware of the relevant identity fact, they link or merge the information about the man on the train and the man on the tele-vision. Most relevantly, this means that, contrary to what we thought, C2 is actu-ally compatible with ascribing communicative success to Loar-case 3. When Jones interprets Smith as intending to refer to the man on the train, one could argue that he thinks a thought with the same truth-conditions as Smith and that he has also recognized Smith's communicative intentions, since, after undergoing the mechanisms of linking and merging, thinking of the relevant man as *the man on the train* or as *the man on the television* are equally correct ways as far as Smith's communicative intentions are concerned. I expect those brief consider-ations to pave the way for an explanation of Unnsteinsson's *data* without letting go of C2.

5. Final Remarks

We started with the simplest criterion for successful communication (C1) and argued that it does not output the correct predictions for Loar-cases. I then assessed Loar's MOP-based account of the problem in light of Buchanan's critical remarks. My conclusion was sympathetic to Buchanan's with the exception of my point that, even if an intentions-based account of communicative success (C2) is, as far as this limited set of data is concerned, less theoretically committal, a MOP-based account is still compatible with everything we have seen so far.

In the following section of the article I then went on to argue against Unnsteinsson's recent suggestions to the effect that the failure of communication

in Loar-cases is to be explained by the presence of contextually relevant false distinctness beliefs about the object being referred to by the speaker. In order to achieve that conclusion, I argued that there can be communicative success even if speaker and hearer hold contextually relevant false distinctness beliefs. I then tried to develop Unnsteinsson's suggestion in a different way but argued that it was bound to collapse into the claim that false distinctness beliefs only disrupt communication when the hearer fails to grasp the speaker's intention – and that this would make his suggestion theoretically redundant. I then finished this section by motivating the claim that, even if speaker and hearer are enlightened, it is not sufficient for successful communication that the hearer entertains a thought with the same truth-conditions as the speaker.

Finally, I turned my attention to Unnsteinsson's remaining challenge: how would a view like C2 account for the fact that we are disposed to ascribe communicative success to cases, like Loar-case 3, where the hearer seems to fail to grasp the speaker's communicative intentions? My reply, admittedly tentative, employed the notions of linking and merging in order to argue that, when thinkers discover that they were acquiring information about a single object by means of two distinct sources, they tend to link or merge this information, and that something like this could very well explain what is going on in the cases to which Unnsteinsson calls attention.

Naturally, there are plenty of open questions we did not have time to touch upon. One of these is: can successful communication be lucky? Just as a hearer might luckily manage to get the truth-conditions of the speaker's thought right, she might also luckily manage to recognize the speaker's communicative intentions. That would mean that C2 is – as much as C1 – subject to cases where a hearer complies with it by means of sheer luck. Then, establishing whether C2 is an adequate criterion for successful communication at all takes us back to the question of whether communicative success can go hand in hand with luck – a question I have not even pretended to be able to answer.

A distinct challenge to C2 draws upon the fact that, as is common for Gricean-inspired views, the resulting account might end up being psychologically too demanding. It is a well-known fact that thinkers usually have different perspectives on the subject matters they think about. Likewise, it is often vague which inferential route to the referent is intended by particular referential speech acts (Peet, 2017). Requiring that the hearer thinks of the referent in exactly the same way as the speaker, or that the hearer magically discovers the exact inferential route hidden behind the speaker's utterance, might quickly lead one to a criterion which is rarely, if ever, satisfied in the real world.

Regardless of the amount of work ahead of us, I expect to have at least made a good case to the effect that progress on the issue of successful referential

communication should not give a role of fundamental importance to the presence or absence of false distinctness beliefs, and that it should drive off the idea that, even for some restricted cases, identity of referential content is enough. Communicating well, just like understanding well and holding the same belief, seems to be a hyper-intensional notion that requires more than an actual match of referents. How exactly we should develop that thought, however, remains a question for future investigation.

Acknowledgements

References

BACH, K. and HARNISH, R. M. (1979) *Linguistic Communication and Speech Acts*. Cambridge, MA: MIT Press.

BUCHANAN, R. (2014) "Reference, Understanding, and Communication." *Australasian Journal of Philosophy* 92(1): 55–70.

BYRNE, A. and THAU, M. (1996) "In Defense of the Hybrid View." *Mind* 105: 139–149.

EVANS, G. (1982) In J. McDowell (ed.), *The Varieties of Reference*. Oxford: Clarendon Press.

GARCÍA-CARPINTERO, M. (2016) "Token-Reflexive Presuppositions and the De Se." In S. Torre and M. García-Carpintero (eds), *About Oneself: De Se Thought and Communication*, pp. 179–199. Oxford: Oxford University Press.

HECK, R. (1995) "The Sense of Communication." *Mind* 104: 79–106.

HYMAN, J. (2012) "Depiction." *Royal Institute of Philosophy Supplement* 71: 129–150.

LOAR, B. (1976) "The Semantics of Singular Terms." *Philosophical Studies* 30(6): 353–377.

PAGIN, P. (2008) "What Is Communicative Success?" *Canadian Journal of Philosophy* 38(1): 85–115.

PAGIN, P. (forthcoming) "When Does Communication Succeed? The Case of General Terms." In T. Marques and Å. Wikforss (eds), *Shifting Concepts: The Philosophy and Psychology of Conceptual Variability*. Oxford: Oxford University Press.

PAUL, M. (1999) *Success in Referential Communication*. Dordrecht: Kluwer.

PEET, A. (2017) "Referential Intentions and Communicative Luck." *Australasian Journal of Philosophy* 95(2): 379–384.

RECANATI, F. (2012) *Mental Files*. Oxford: Oxford University Press.

RECANATI, F. (2016) "Indexical Thought: The Communication Problem." In S. Torre and M. García-Carpintero (eds), *About Oneself: De Se Thought and Communication*, pp. 141–178. Oxford: Oxford University Press.

SALMON, N. (1986) *Frege's Puzzle*. Cambridge, MA: MIT Press.

SCHIFFER, S. (1972) *Meaning*. Oxford: Clarendon Press.

STRAWSON, P. (1974) *Subject and Predicate in Logic and Grammar*. London: Methuen.

UNNSTEINSSON, E. (2016) "Confusion Is Corruptive Belief in False Identity." *Canadian Journal of Philosophy* 46(2): 204–227.

UNNSTEINSSON, E. (2018a) "Referential Intentions: A Response to Buchanan and Peet." *Australasian Journal of Philosophy* 96(3): 610–615.

UNNSTEINSSON, E. (2018b) "The Edenic Theory of Reference." *Inquiry: An Interdisciplinary Journal of Philosophy* 62(3): 276–308.