

# On testing for de se and de re construals across languages \*

Hazel Pearson  
Queen Mary University of London

**Abstract:** Putatively de se expressions such as shifted indexicals and logophoric pronouns have become an object of considerable interest among formal semanticists and philosophers of language over the past 20 years. Fieldwork on understudied languages such as Amharic, Zazaki, Nez Perce, Ewe and a number of sign languages has played a crucial role in this research, resulting in a large inventory of pronouns and anaphors that have been described as unambiguously de se. Yet the question of how to go about eliciting data from a native speaker consultant about whether some pronoun or anaphor is obligatorily read de se has seldom been explicitly discussed. In Pearson (2013, 2015) I argued that the claim that logophoric pronouns are unambiguously de se is incorrect, at least for the Niger-Congo language Ewe. In this paper, I discuss the methodological difficulties that arise with elicitation work in this domain and describe how I have approached them in my research. There are two main issues: (i) the unusual or implausible nature of the ‘mistaken identity’ scenarios that must be presented to the consultant and (ii) the tendency of mistaken identity scenarios to systematically induce a bias against the de re reading. The difficulty presented by the first issue has been overestimated, while the second one has to my knowledge not been discussed before. I propose that (ii) is related to a broader observation about truth value judgment tasks: when asked to give a truth value judgement for a sentence S relative to a scenario, linguistically naïve consultants’ judgments are likely to be conditioned by an assumption that every feature of that scenario is relevant. I argue that this tendency leads to a bias in favour of de se over de re construals and propose some strategies for overcoming the difficulty that this presents for diagnosing de re readings.

**Keywords:** de se, attitude reports, indexical shift, logophoric pronouns, Ewe

## 1 Introduction

In 2007, the actor Ian McKellen played the role of King Lear – a role for which he was required to grow a beard. On being complimented on the beard in a television interview, McKellen responded, ‘It’s taken a year to grow, I hate it. I walk along the street and see this geezer in shop windows and I think, “Oh dear, I’m glad I’m not him!”’<sup>1</sup>

Here McKellen is of course having fun with the idea of a change in physical appearance that could cause someone to fail to recognise himself. When that happens an agent can end up thinking about himself in the third person – the ‘him’ who McKellen sees in a shop window, who is in fact his own reflection. In that moment, the agent is *unaware* that he is thinking about himself. This explains the use of the third person pronoun; when an agent knows that he is thinking (or talking,

---

\* My thinking on the topics addressed in this paper has been informed over the years by conversations with Abigail Bimpeh, Gennaro Chierchia, Amy Rose Deal, Tom Grano, Irene Heim, Dan Lassiter, Tom Roeper, Frank Sode, among others. Earlier versions of this work were presented at ZAS Berlin, UMass Amherst and the Workshop on Pronouns in Embedded Contexts at the Syntax-Semantics Interface. I also thank Seth Cable, Jérémy Pasquereau and an anonymous reviewer for insightful comments that led to a much improved paper.

Contact info: h.pearson@qmul.ac.uk

<sup>1</sup>Parkinson (2007), <https://www.youtube.com/watch?v=UCytMMGmxBo>

etc.) about himself, he instead uses *I*. We thus have two possible perspectives upon the self – a first personal one, and a third personal one.

Beliefs that involve thinking about oneself in a first person way are called *de se* beliefs. A *de re* belief is a belief that is about some individual, who may but need not be the belief holder. Unlike with *de se* beliefs, a *de re* belief that is about the belief holder is neutral with respect to the perspective that the belief holder takes upon herself; it does not matter whether the perspective on the self is first personal or third personal. *De se* beliefs are thus subject to a condition of *awareness* that does not apply with *de re* beliefs. So all *de se* beliefs are also *de re* beliefs, but not vice versa. *Merely de re* beliefs – beliefs that are about the attitude holder but are not *de se* - can be recognised by setting up ‘mistaken identity’ scenarios involving predicaments such as McKellen’s.

A striking fact about natural language is that it appears to be sensitive to the distinction between *de se* and merely *de re* beliefs. In particular, it has been claimed that there are (i) belief reports that can be used to ascribe both *de se* and merely *de re* beliefs, and (ii) belief reports that obligatorily describe *de se* beliefs. (Note that since all *de se* beliefs are also *de re* beliefs, reports that belong to the second class entail the corresponding *de re* belief ascriptions.)

Take, for instance, the following sentence.

(1) McKellen believed that the man with a beard looked terrible.

In a scenario where McKellen sees a man with a beard and thinks ‘that guy looks terrible’, (1) is clearly true. McKellen’s anecdote furnishes us with just such a scenario. Of course, in this scenario we have two further details: (i) the man with the beard is McKellen himself, and (ii) McKellen is not aware of this. So now let’s ask whether in the scenario of McKellen’s anecdote, (2) is true.<sup>2</sup>

(2) McKellen<sub>i</sub> believed that he<sub>i</sub> looked terrible.

On the one hand, McKellen clearly does not look at the man with the beard and think, ‘I look terrible’; his failure to recognise himself is after all the source of the joke. On this basis, one might be inclined to judge the sentence *false*. On the other hand, consider the following premises: (i) McKellen believed that the man with a beard looked terrible; (ii) McKellen is the man with a beard. Taken together, these premises would seem to license the conclusion that (2) is *true* in the scenario of McKellen’s anecdote. It seems then that (2) is ambiguous between a reading on which it reports a *de se* belief, and a weaker reading on which it reports a *de re* belief. On the former reading – the *de se* reading - the sentence is false, and on the latter one – the *de re* reading - it is true.

(2) involves the ordinary third person pronoun *he* – a pronoun that can but need not refer to the subject of *believe*. (In a context that makes salient some other suitable individual, the pronoun could refer to them.) But across languages, we find various pronominal and anaphoric expressions whose referential properties are more constrained. Take for instance PRO, the unpronounced element that occurs in the embedded subject position of obligatory control sentences. English *believe* is not a control verb, but doxastic predicates in many other languages are. For example, (3) is a German counterpart of (2) with the control predicate *glauben*.

(3) McKellen glaubt PRO furchtbar auszusehen.  
McKellen believes PRO terrible look  
Lit: ‘McKellen<sub>i</sub> believes PRO<sub>i</sub> to look terrible.’

<sup>2</sup>I use coindexation to indicate intended coreference; the indices are not meant to have theoretical status.

In the scenario of McKellen's anecdote, (3) is unambiguously false. In general, attitude reports formed with control predicates, with an infinitival complement with PRO as the subject are said to unambiguously report *de se* attitudes (Morgan 1970; Chierchia 1990). We thus have a contrast between ordinary third person pronouns, and PRO: the former figures in attitude reports that are ambiguous between *de se* and *de re*, while attitude reports involving the latter are unambiguously read *de se*. We will say that a pronoun or anaphor is ambiguous between a *de se* and a *de re* reading if it occurs in attitude reports that show *de se/de re* ambiguity. If attitude reports in which it occurs are unambiguously read *de se*, then we shall say that the pronoun/anaphor itself is unambiguously (or obligatorily) *de se*.

The literature abounds with examples of pronominal and anaphoric expressions that at first glance seem more like overt counterparts of PRO than ordinary pronouns. These expressions include logophoric pronouns of the type attested in West African languages, shifted indexicals, long-distance reflexives, and conjunct-disjunct marking. These expressions have in common that when embedded below an attitude verb, their interpretation appears to be anaphorically dependent on some designated argument of the embedding predicate – typically, the person whose attitude is being reported. It is tempting, then, to assume that like PRO, these expressions will be unambiguously interpreted *de se*. So if we find a language with such an expression  $\alpha$ , we might formulate the hypothesis that  $\alpha$  is obligatorily interpreted *de se*.

This paper is concerned with the question of how to test this hypothesis in semantic fieldwork on an unfamiliar language. I will make the following two claims: (i) the literature is full of examples of cases where the hypothesis that a given expression  $\alpha$  in a given language is obligatorily *de se* is taken to be true; and yet (ii) the scenarios that are typically used to test for *de se/de re* ambiguity systematically bias the consultant towards *de se* readings, and hence are not well suited to the task of uncovering *de re* readings. In fact, I'll go so far as to say that we currently lack a robust, agreed upon methodology for diagnosing *de re* readings. I will make some modest suggestions aimed towards improving this state of affairs.

This paper is structured as follows. In Section 2, I discuss in greater detail why it is of theoretical interest to establish which pronouns and anaphors are unambiguously read *de se*. In Section 3 I present as a case study my own work on the logophoric pronoun in the Niger-Congo language Ewe (Pearson 2013, 2015). In these studies, I showed that contrary to the received wisdom of the time, this element does in fact have a *de re* reading. I describe the elicitation method that I followed in this work. In so doing, we encounter a problem that lies at the heart of empirical work in this domain: what is typically considered evidence that a pronoun/anaphor  $\alpha$  is unambiguously read *de se* is also compatible with it being ambiguous between a *de se* and a *de re* reading, with judgments reflecting not the absence of the latter, but rather a preference for the former relative to the presented contexts. Sections 4 and 5 are concerned with addressing this problem: I propose three strategies that can be used to (attempt to) falsify the hypothesis that  $\alpha$  is obligatorily *de se*. Section 6 considers as an alternative the possibility that rather than pursuing these strategies, it may be sufficient to simply compare the target sentence to a suitable baseline. According to this view (implicit in much earlier work), if a contrast is perceived for the resulting minimal pair, this may be sufficient evidence that  $\alpha$  is obligatorily *de se*; I argue that this assumption is incorrect. Section 7 concludes.

## 2 Motivations

Let me begin by considering *why* one might want to know whether an expression  $\alpha$  is unambiguously *de se*. After all, the distinction between *de se* and *de re* appears to be relevant only in ‘mistaken identity’ scenarios, which are very rare in ordinary life; when we think about ourselves, we usually know that we are doing so.

One could of course get by as a non-native German speaker, say, without knowing that when an attitude verb like *believe* takes an infinitival complement, the resulting sentence is obligatorily read *de se*. The unusual nature of mistaken identity scenarios means that one is very unlikely to encounter in ordinary conversation a misunderstanding that pivots on the choice between a *de se* and a *de re* construal of a pronoun or anaphor. But I shall assume that the goal of the semantic fieldworker is not (primarily) to construct pedagogical materials for second language learners, but rather to develop a descriptively adequate account of the portion of the grammar that she is interested in. If for instance one is studying an expression  $\alpha$  which can (or must) occur in the scope of an attitude predicate and can (or must) correspond to the attitude holder, then one is in the business of discovering what the truth conditions of attitude reports containing  $\alpha$  are; whether or not  $\alpha$  is unambiguously *de se* is an important facet of such a project.

A project of this type also has broader implications for inquiry into linguistic meaning. One question underpinning much related research over the past two decades or so is, ‘How extensive is the inventory of unambiguously *de se* expressions that we find in natural language?’ (e.g. Anand 2006; Deal 2020; Schlenker 1999, 2003). If there should turn out to be a large inventory of expressions of this type, this would provide evidence that the grammar ‘cares’ about the *de se/de re* distinction – that is, that the distinctive psychological class of mental attitudes that we call *de se attitudes* corresponds to a distinctive grammatical class of attitude reports that we can call *de se reports*. This would be an intriguing result precisely because it seems that natural language users should be able to get along perfectly well without a dedicated means of reporting *de se* attitudes: after all, in a scenario where a *de se* report is true, the corresponding *de re* report is also true (at least in positive environments). While the reverse does not hold, the scenarios where the two classes of report differ in truth value – mistaken identity scenarios – are in practice extremely rare. Why then should the grammar take the trouble to code *de se*?

A related goal is to choose between competing theories of *de se* construal. One camp holds that there is a dedicated grammatical mechanism for generating *de se* truth conditions (e.g. Anand 2006; Chierchia 1990; Patel-Grosz 2020; Pearson 2013, 2018; Percus and Sauerland 2003). Typically, such approaches take it that the proper treatment of *de se* requires a departure from the classical Hintikka analysis of attitude verbs (Hintikka 1969): for proponents of this view, attitude predicates are not quantifiers over worlds, but over objects that are more fine-grained – treated variously as ‘centred worlds’ or ‘contexts’. Another view holds that *de se* can simply be treated as a ‘special case’ of *de re*; since an account of *de re* is needed anyway, an account of *de se* can be provided without any additional theoretical machinery (Landau 2015, 2018; Maier 2006, 2009, 2011; Reinhart 1990; Santorio 2014). Proponents of the first view generally assume that *de se-as-de re* accounts will struggle to cope with unambiguously *de se* expressions, since under such a view it is not clear how to rule out the *de re* reading. So if there should turn out to be robust cross-linguistic evidence for an extensive inventory of pronouns and anaphors that are unambiguously read *de se*, this would seem to be a compelling argument that there is a dedicated grammatical mechanism underlying *de*

se interpretations.<sup>3</sup>

Reading the literature on logophoric pronouns, shifted indexicals and the like, one is left with the impression that the results are already in: in (4) I provide a sample of the expressions that have at some time or another been claimed to be unambiguously *de se*.

- (4) a. Logophoric pronouns<sup>4</sup>  
e.g. Bafut (Kusumoto 1998), Tangale (Haida 2009)
- b. Shifty indexicals<sup>5,6</sup>  
e.g. Amharic (Schlenker 1999), American Sign Language (Schlenker 2017), Korean (Park 2018), Mishar Tatar (Podobryaev 2014), Nez Perce (Deal 2020), Uyghur (Sudo 2016), Zazaki (Anand 2006)
- c. Long-distance reflexives<sup>7</sup>  
e.g. Korean (Park 2018), Chinese (Huang and Liu 2001), Japanese (Oshima 2004)
- d. Conjunct-disjunct systems<sup>8</sup>  
e.g. Newari (Zu 2018)

Here I'll pause to remark that the claim that shifted indexicals in Amharic are unambiguously *de se* deserves particular scrutiny. Philippe Schlenker's PhD dissertation (Schlenker 1999) is typically cited as the original source for this claim. Here is his example:

- (5) Scenario: John, who is a candidate in the election, is so drunk that he doesn't remember who he is. He watches TV and sees a candidate he finds terrific, thinking that this guy must be a hero. This candidate happens to be John himself, though he doesn't realize it.  
John əne Jägna näNN alä  
John I hero am said  
'John<sub>i</sub> said that he<sub>i</sub> is a hero.'  
[Schlenker 1999: 97, ex (17b)]

Schlenker reports that his consultant judged the sentence false relative to the presented situation. Moreover, this contrasts with a baseline sentence where the shifted indexical is replaced by a definite description meaning 'the man': (6) is judged true in the scenario in question.

<sup>3</sup>This landscape is changing, however, with work by Landau (2015, 2018), Maier (2011), Santorio (2014) and Schlenker (2003) appealing to a *de se-as-de re* analysis to account for the obligatorily *de se* nature of PRO.

<sup>4</sup>A logophoric pronoun is a pronoun that obligatorily occurs below an attitude predicate and obligatorily corresponds to the attitude holder.

<sup>5</sup>A shifty indexical is an indexical whose interpretation can shift in certain environments (canonically, the complement of attitude verbs), so that it does not have to pick up its referent from the context of utterance. For example, in a language where the first-person pronoun can shift, *I* in 'John thinks that I am a genius' may correspond to John rather than the utterance context speaker.

<sup>6</sup>This list concerns only shifty 1st person indexicals. Related questions concerning 'de te' interpretations of shifted second person indexicals and 'de hic' interpretations of expressions like *here* are set aside.

<sup>7</sup>A long-distance reflexive is a reflexive that can take as its antecedent an argument in a higher clause.

<sup>8</sup>Conjunct-disjunct systems employ dedicated morphology that when used in the clausal complement of an attitude predicate imposes the requirement that the embedded subject must correspond to the attitude holder.

- (6) John səwyewJägna näw alä  
 John the-man hero is  
 ‘John said that the man is a hero.’ [Schlenker 1999: 97, ex (17a)]

Schlenker concludes that shifted indexicals are indeed obligatorily *de se* in Amharic. Yet Malamud (2006) presents data suggesting that Amharic shifted indexicals can have *de re* readings after all. Here is her example.

- (7) Scenario: A certain class has no T.A., so Hans must do both the lecturer’s and the assistant’s work, like grading homeworks. Hans goes to a happy hour one evening, where all the professors and T.A.s are gathering. He gets drunk, forgets that he himself is the grader, and asks someone, who is the grader for my course? His interlocutor, being tongue-in-cheek, points to Hans’s reflection in the bar mirror, and says, that guy. Hans is very drunk, and he doesn’t realize that his interlocutor pointed to a reflection. He says “The comments on all of the homeworks are very good and detailed this term – this guy, my course assistant, works very hard!”  
 profäsəru bät’am bəzu səra ə-sär-allähu alä  
 professor very much work 1S-WORK.IMP-AUX 1S SAY.PRF.3SM  
 ‘The professor said that he works very hard.’ [Malamud 2006: 104, ex (187)]

Malamud reports that her informants ‘show no hesitation’ in judging the sentence true in this scenario (106).

What should we make of these seemingly conflicting empirical claims? Notice that the first-person pronoun is pronounced in Schlenker’s example but not in Malamud’s. So one possibility is that overt shifted indexicals are obligatorily read *de se*, but that in cases of pro-drop, a *de re* reading is possible. This hypothesis would certainly be interesting if true<sup>9</sup>, but to test it, we need minimal pairs: sentences differing only with respect to whether the shifted indexical is overt or covert, evaluated with respect to the same scenario.

Moreover, Schlenker appears to have undertaken further consultant work after submission of the dissertation, subsequently leading him to be more cautious in his appraisal of the data: in the revised version of the dissertation he presents the same scenario and pair of sentences but writes, ‘Unfortunately, my informant finds the sentences almost impossible to assess (although it is not clear why)... I leave the result as a prediction to be tested and challenged rather than as a hard fact’ (Schlenker 2000: 126).<sup>10</sup>

The claim that Amharic indexicals are obligatorily *de se* has often been repeated and has provided impetus for a wealth of cross-linguistic research into the alleged connection between indexical shift and *de se*. Yet as this brief discussion highlights, it is far from clear what the facts really are even for as famous a case as Amharic. In this paper, I want to outline some methodological considerations that I hope can inform future fieldwork on shifted indexicals in Amharic and related phenomena in other languages. I will discuss the following general problem: to claim that an expression  $\alpha$  is

<sup>9</sup>It would contrast, for instance, with findings on Kutchi Gujarati reported in Patel-Grosz (2020). This language does not have indexical shift, but it does have pro-drop; Patel-Grosz claims that overt pronouns are read *de re*, whereas null pronouns are obligatorily *de se*. She connects this finding to a proposed correlation between strong pronouns and *de re* readings, and between weak pronouns and *de se* readings; if the above hypothesis for Amharic turned out to be true, this would seem to challenge that proposal.

<sup>10</sup>The discrepancy between the two versions of Schlenker’s dissertation was uncovered over the course of an email exchange with Tom Grano; I am grateful to him for helping me to untangle the historical record here.

unambiguously de se is to presuppose that the researcher has a reliable method for testing the hypothesis that  $\alpha$  must be read de se and cannot be read de re.<sup>11</sup> In this paper, I'll call this assumption into question by arguing that the contexts that are typically used to elicit judgments on de se/de re readings tend to bias consultants towards the de se reading, resulting in a risk of underdiagnosing de re readings. To make matters worse, I'll suggest that among those expressions that are ambiguous between de se and de re, the preference for de se readings is stronger for some pronouns than others (e.g. English *he himself*). If  $\alpha$  happens to trigger a particularly strong de se preference, then the risk of underdiagnosing de re readings is even greater. In the next section, I turn to my own work on the logophoric pronoun in Ewe as a starting point for discussion of these issues.

### 3 The logophoric pronoun in Ewe

Logophoric pronouns of the kind commonly found in West African languages are elements that (i) obligatorily occur in the scope of an attitude and (ii) must correspond to the attitude holder.<sup>12</sup> The examples below, from the Niger-Congo language Ewe, illustrate this point.

- (8) Kofi be yè dzo  
 Kofi say LOG leave  
 'Kofi<sub>i</sub> said that he<sub>i</sub>/<sub>\*j</sub> left.'

- (9) \*yè dzo  
 LOG leave

[Pearson 2015: 78, exs (1), (2)]

Until about ten years ago, the Ewe logophoric pronoun *yè* was considered the poster child example of an unambiguously de se element – its distribution was taken to suggest that it was a kind of 'overt PRO' (Schlenker 1999, 2003, 2011; von Stechow 2003; Heim 2002). But at that time, the fieldwork had not actually been done to establish whether *yè* is indeed obligatorily read de se (as various authors acknowledged to varying degrees). In elicitation work with 5 native speakers, I found that surprisingly, *yè* does in fact have a de re reading. This finding has since been replicated in subsequent studies (O'Neill 2016; Bimpeh 2021; Satik 2019), though see Bimpeh (2019) for a dissenting view.

My speakers were five bilingual Ewe/French speakers. I worked with one of these speakers in face-to-face elicitation sessions conducted in Cambridge, Massachusetts every two weeks over a period of a year. Elicitation with the other four speakers took place over Skype, meeting for a smaller number (between two and eight) of one-hour sessions. Just one of the five speakers was still living in Togo at the time.

Semantic judgments were collected according to the following procedure. First the target sentence was constructed and checked with the consultant for acceptability. Then, a scenario was presented to the consultant. For the elicitation that was conducted over Skype, a typed out version of the

<sup>11</sup>Here I am concerned primarily with empirical investigation of a language which the researcher does not speak natively herself.

<sup>12</sup>There is some controversy over claim (i), since logophoric pronouns can often occur in purpose clauses (including in Ewe (Bimpeh 2021: 53)). For present purposes I will assume that the generalization is essentially correct, and that purpose clauses involve a covert attitudinal operator with meaning akin to 'intend'.

scenario was sent to the consultant as an instant message. Next, I allowed time for the consultant to read the scenario (if appropriate) and asked whether it was clear and whether they had any questions. Once I was confident that the consultant had understood the scenario, I presented the target sentence and asked whether they thought that it was true or false in the scenario. Finally, I encouraged the consultant to give reasons for their answer and to add any further comments.

Here is an example of the data that was elicited following this method.

- (10) Scenario: John has just found an old paper that he wrote, but he doesn't realize that he is the author of the paper. He reads it and is impressed by what a good paper it is. He says, "Whoever wrote this paper is clever".

John be yè le cleva

John say LOG COP clever

'John said that he was clever.'

[Pearson 2015: 80, ex (9)]

Four of the five speakers judged the sentence to be true in this scenario. (In a moment I will consider what we should make of the judgments of consultant 3, for whom the sentence was false.) I also elicited judgments relative to a real-life example of mistaken identity found in a news story:

- (11) Scenario: Following a spate of burglaries, a policeman was alerted by CCTV operators that someone was acting suspiciously, and chased after that individual, unaware that it was himself.

Sodza xose be yè nyi fianfito.

Policeman believe COMPL LOG COP thief

'The policeman<sub>i</sub> believed that he<sub>i</sub> was a thief.'

[Pearson 2015: 99, ex (53b)]

For this example, I had consultants read the newspaper story themselves.<sup>13</sup> Again, four of the five speakers judged the sentence true relative to this scenario.

Based on these, and similar data, I concluded that the logophoric pronoun in Ewe has a *de re* reading, contrary to earlier claims in the literature. But what should we make of the speaker who judged the sentences *false*? There are three possible interpretations of the data that I can see: (i) this speaker's grammar is different from that of the other four: *yè* is obligatorily *de se* in the grammar of that speaker; (ii) the speaker's grammar is the same as that of the other four: *yè* is obligatorily *de se* in the grammar of all five speakers. This would mean that for the other four speakers, the fact that they judged the sentences true in mistaken identity scenarios does not indicate a *de re* reading for *yè*, and must be explained some other way; (iii) the fifth speaker's grammar is the same as that of the other four: *yè* has a *de re* reading for all five speakers, but the fifth speaker has a preference for interpreting pronouns that show *de se/de re* ambiguity relative to the *de se* reading in the presented contexts. I'll now explain why I find interpretations (i) and (ii) unlikely, and why I am persuaded by interpretation (iii).

Interpretation (i) assumes that there can be inter-speaker variation in whether or not a logophor can receive a *de re* reading. This would mean in effect that there are two grammars of Ewe. In grammar 1, *yè* is unambiguously read *de se*. In grammar 2, *yè* is ambiguous between a *de se* and a *de re* reading. But this raises the question of how the child would converge on one of these grammars.

<sup>13</sup>The article can be found here: <https://uk.news.yahoo.com/policeman-'chased-himself'-for-20-minutes-while-looking-for-suspect.html?guccounter=1>. I first learned about this case from a social media post by Rob Truswell.



After all, the input to acquisition is highly unlikely to contain data that would settle the matter directly: the mistaken identity scenarios where the *de se/de re* distinction makes its presence felt are very rare in ordinary life. Even if the child were to encounter such a scenario, she would presumably lack the negative evidence needed to determine that a given logophor does not have a *de re* reading; caregivers would be very unlikely to mention that the logophor is inappropriate in the situation, and even if they did, the child would be unlikely to be attentive to that information. Perhaps a more likely scenario is that the two grammars would differ in terms of some feature or property that serves as a cue for the availability (or lack thereof) of the *de re* reading. I was unable to identify a difference in the idiolects of the consultants that would point to this interpretation, although this does not of course exclude the possibility that such a difference may yet be discovered.

According to interpretation (ii), the speakers who judged the critical sentences true in mistaken identity scenarios do not in fact have *de re* readings for  $yè$ . This interpretation would require an explanation for the judgments that the speakers gave; one possibility is that they misunderstood the scenarios that were presented to them. Indeed, one sometimes finds in the literature mention of the problem of mistaken identity scenarios being esoteric or difficult to understand, and claims that this consideration may cast doubt on the judgments obtained (e.g. Bimpeh 2019).

Here I think we should give more credit to our consultants. Speakers who are not trained linguists are certainly capable of understanding mistaken identity scenarios, and, with a little practice, to introspect about the truth values of sentences relative to such scenarios.<sup>14</sup> In the case of the consultants I was working with, it was helpful to conduct the work over an extended period of time, with regular meetings allowing time for consultants to get used to the elicitation process and gain practice in reasoning about the truth or falsity of sentences in particular scenarios. The use of the policeman incident provided a naturalistic example accessible to anyone able to read a newspaper story. Over time, consultants became familiar with the general shape of mistaken identity scenarios, and they showed recognition and amusement when new ones were presented. They were given opportunities to ask for clarification if anything was unclear, and judgments were not elicited until I was confident that they were secure in their understanding of the scenario.

That leaves interpretation (iii):  $yè$  has a *de re* reading in the grammar of all five speakers, but consultant 3's answers reflect a preference for interpreting sentences relative to the *de se* reading in the presented contexts. It is not uncommon for speakers to show such a preference: many speakers will interpret (2), for instance (repeated below) as false relative to the scenario that this paper began with.

(12) McKellen<sub>i</sub> believed that he<sub>i</sub> looked terrible.

Ask any group of native English speakers whether this sentence is true, and some will respond that it is false, because McKellen didn't think 'I look terrible'. This is consistent with *he* having a *de re* reading: the sentence *also* has a *de se* reading, and it is false relative to that reading. So consultant 3's answers are compatible with him having a *de re* reading for  $yè$ , but interpreting the target sentences relative to the *de se* reading in the presented contexts. Pearson (2015) reasoned that if this is the correct interpretation of the elicited judgments, then the preference for *de se* readings

---

<sup>14</sup>I do not of course mean to suggest that consultants never experience any difficulty at all with truth value judgment tasks in this domain; Schlenker's remarks from the revised version of his dissertation suggest that his consultant certainly did, and Bimpeh (2019) also reports confusion on the part of her consultants. I merely want to point out that such difficulties are not insurmountable.

should also show up with ordinary pronouns embedded in attitude reports in French (which the consultant also speaks natively). Sure enough, the consultant judged (13) false in the mistaken identity scenario.

- (13) John croit qu'il est voleur.  
John believe that he is  
'John believes that he is a thief.'

[Pearson 2015: 100, ex (55)]

It seems right then to conclude that *yè* has a de re reading. The discussion raises three further questions. Firstly, if speakers generally show a strong preference for de se readings, why did four of the five consultants readily interpret the critical sentences on their de re readings? I will leave this as an open question, other than to speculate that it may have simply been a statistical fluke. In my view, though, the tendency to prefer de se readings makes the cases that we find of de re readings all the more compelling: the odds are stacked *against* finding de re readings, and yet the majority of our consultants detected these readings anyway. Secondly, where does the apparent preference for de se readings come from? And finally, given such a preference, how can the absence of a de re reading ever be demonstrated? That is, if an attitude report containing the pronoun/anaphor of interest is judged false in a mistaken identity scenario, how can we rule out the possibility that the expression has a de re reading, but the speaker merely interpreted the sentence with respect to the de se reading? I turn to the second question in section 4, and the third one is addressed in sections 4 and 5.

#### 4 Explaining, and overcoming, the de se preference

Let's put ourselves in the position of the native speaker consultant for a moment. When presented with a scenario and asked for a truth value judgment, it seems reasonable to assume that all of the information included in the presented scenario is somehow relevant: since elicitation work is a cooperative activity, the fieldworker presenting the scenario can be presumed to be observing Grice's maxim of relevance (Grice 1970). I will call this assumption *Modified Chekov's Gun*.<sup>15</sup>

For any feature F of a scenario, Modified Chekov's Gun gives rise to a *presumption of relevance* for F. One way (but not the only way) that F can be relevant in the context of the elicitation session is if it plays a role in the reasoning that leads to the requested truth value judgment. So a consultant can maintain the presumption of relevance for F by assuming that the fieldworker intends for F to figure in that reasoning. This has consequences for what happens when a consultant is asked to provide a truth value judgment about a sentence that is ambiguous between a reading A where F serves as grounds for the choice between a judgment of 'true' or 'false', and a reading B where F provides no such grounds: in such a predicament, she can preserve the presumption of relevance for F by interpreting the sentence with respect to reading A. These are the circumstances that a consultant is typically in when she is asked to give a judgment about a (potentially) de se/de re ambiguous sentence.

---

<sup>15</sup>Chekov's Gun is the name given to the principle, credited to the playwright Anton Chekov, that every element that is introduced into a story – such as a gun placed on the stage, for instance – should play a role in that story. Note that Chekov's Gun is a principle to be followed by writers, whereas Modified Chekov's Gun is an assumption on the part of native speaker consultants.

As we have seen, the scenarios that are apt for testing for de se/de re ambiguity involve ‘mistaken identity’ about the attitude holder: the attitude holder fails to recognise herself in some way. Recall also that de se readings are strictly stronger than de re readings: the truth of a de se belief report entails that the attitude holder holds the corresponding de re belief. But in a mistaken identity scenario, a de se/de re ambiguous sentence can be true on its de re reading but false on its de se reading. In such a context, if the target sentence is interpreted with respect to a de se reading, then it is judged false *because of* the mistaken identity. But if it is interpreted on a de re reading, the mistaken identity is *irrelevant* to the truth of the target sentence. (If for instance McKellen suddenly realises that he is looking at his own reflection, and thinks ‘*I look terrible!*’, then ‘McKellen believes that he looks terrible’ is still true on its de re reading.) In the case of de se readings, the mistaken identity serves as grounds for the judgment; with de re, the mistaken identity is irrelevant for the purposes of the judgment. So interpreting the sentence with respect to its de se reading provides a means of preserving the presumption of relevance for the mistaken identity feature; interpreting it on its de re reading does not. All else being equal, the speaker performing this task will therefore be more likely to interpret a de se/de re ambiguous sentence with respect to its de se reading. I propose that this is what underlies what I will refer to as the ‘de se bias’. From this perspective, there is no *absolute* preference for de se readings: it is not a property of de se readings *per se* that they are preferred to de re ones. Rather, it is a property of the contexts that are typically used in elicitation that they tend to bias consultants towards such readings.

The good news for the researcher who wants to avoid underdiagnosing de re readings is that what is relevant in a discourse context can shift as the discourse proceeds. How this unfolds depends on the evolving communicative goals of the interlocutors, which I take to be reflected in the Question under Discussion that is current at a given point in the conversation (Roberts 2012). If the account sketched above is correct, it should be possible to overcome the de se bias by embedding the target sentence in a discourse context where there is no expectation that the mistaken identity feature of the scenario will be treated as relevant to the interpretation of that sentence. I will present two examples from previous work that suggest that this prediction is borne out. In addition to providing evidence that the account just sketched is along the right lines, we shall see that the two cases furnish us with strategies for overcoming the de se bias.

#### 4.1 *Wh*-clarification questions

Suppose that I tell the Ian McKellen story to Salah, who asks for clarification in the following way.

(14) Salah: ‘I don’t understand the story. *Who* did McKellen believe looked terrible?’

I can answer with the sentences in (15), with focal stress on the embedded subject.

- (15) a. McKellen believed that *he* looked terrible.  
 b. McKellen believed that *he himself* looked terrible.  
 c. McKellen believed *himself* to look terrible.

That (15a) is true in the scenario is unsurprising. But the truth of (15b) is more surprising given the longstanding assumption dating back to Geach (1957) that *he himself* is unambiguously de se. Pearson (2017) argued on the basis of data analogous to (15b) that this assumption is false,

concluding that although there appears to be a preference to interpret *he himself* with respect to its de se reading, this preference can be overcome given an appropriate discourse context such as that provided by a *wh*-question. Example (15c) suggests that this is also the appropriate conclusion to draw about ECM reflexives, in contrast to earlier work arguing that they do not allow de re readings (Chierchia 1990, Hornstein and Pietroski 2010, Moulton 2008).<sup>1617</sup>

Salah's question in (14) shifts the discourse to one where the Question under Discussion is 'Who did McKellen think looked terrible?' I assume, following Hamblin (1958) that the denotation of a question is the set of possible answers to that question. The denotation of Salah's question might be:

- (16) { $\lambda w$ . in  $w$ , McKellen believed that McKellen looked terrible,  
 $\lambda w$ . in  $w$ , McKellen believed that Patrick Stewart looked terrible,  
 $\lambda w$ . in  $w$ , McKellen believed that Maggie Smith looked terrible,  
 $\lambda w$ . in  $w$ , McKellen believed that Judi Dench looked terrible, ...}

Given a domain of individuals  $D_e$ , we have:

- (17)  $\llbracket$ Who did McKellen believe looked terrible? $\rrbracket = \{p: \exists u \in D_e [p = \lambda w$ . in  $w$ , McKellen believed that  $u$  looked terrible] $\}$

The set of propositions denoted by Salah's question is thus a set of de re belief attributions: Salah is asking for the *res*, the individual who McKellen's belief is about. For these purposes, the perspective under which McKellen is thinking about that person is irrelevant; indeed the fragment 'McKellen!' would also be a suitable answer. More technically, the question is answered by supplying a witness for the existential quantifier in (17) – the individual who, when plugged in as a value for  $u$  yields a true proposition. A congruent answer to Salah's question will thus be a member of the set of de re belief reports that it denotes. So assuming that the consultant takes the target sentences in (15) to be intended as congruent answers to the QUD, she will interpret them with respect to their de re reading, and therefore judge them true despite the mistaken identity feature of the scenario.<sup>18</sup>

<sup>16</sup>Hornstein and Pietroski would perhaps be unsurprised by the facts reported here: they acknowledge in a footnote that they found that some speakers accept de re readings of ECM reflexives with focal stress (2010: 68 fn.1). Moulton argues that ECM reflexives tolerate de re readings in the scope of *want* and *expect* but not *believe* or *consider*.

<sup>17</sup>There is surely more to say about why the preference for de se readings should be particularly strong with ECM reflexives and *he himself*. Following Pearson (2017), which in turn builds on remarks in Schlenker (2000: 59), it is tempting to relate this to the emphatic nature of the reflexive. Why it should be that it is the *de se* nature of the reported belief that is taken to be emphasized remains open to further investigation, however. Note also that the difficulties encountered with investigating de se/de re ambiguities in other languages will be particularly acute if the expression under investigation in that language should happen to be like ECM reflexives or *he himself* in triggering a particularly strong de se preference: if we struggle to detect the de re reading of these terms even in English, it will be all the more difficult to do so in fieldwork on an unfamiliar language.

<sup>18</sup>If this reasoning is correct, then we might expect that a report formed with an expression that is obligatorily de se will be infelicitous in this type of discourse context: a proposition  $p$  expressed by a de se report is not among the possible answers to the QUD. Of course,  $p$  may *entail* such a possible answer, but it will be over-informative. (cf. Groenendijk and Stokhof 1984, who propose that if  $q$  is a possible answer to some question and  $p$  entails  $q$ , then  $q$  is a better answer than  $p$ .) I leave it to future work to test this hypothesis; it cannot be tested with unpronounced expressions such as PRO, which cannot bear focal stress and therefore yield infelicity in such contexts for independent reasons.

Notice that the mistaken identity feature of the scenario has not been eliminated in this set up. Indeed it should not be, if we want to be able to test for de se/de re ambiguity. But its influence in favouring one reading over another has been overcome by using an explicit question to shift the context to one where mistaken identity is not presumed to be relevant to the immediate communicative goals of the interlocutors. That this strategy is successful in making de re readings available supports the view that the general preference for de se readings is an artefact of the contexts typically used, which tend to trigger the presumption that mistaken identity is relevant.

If this discussion is along the right lines, then it suggests a robust strategy for testing for de re readings that can serve to overcome the de se bias: if in a mistaken identity scenario the target sentence is judged to be a felicitous and true answer to a *wh*-question such as ‘Who did McKellen believe looked terrible?’, this counts as evidence that the expression under investigation is not obligatorily de se. Along the way, we have encountered a further issue for the fieldworker eliciting judgments on de se/de re to keep in mind: there are anaphors that appear to trigger an especially strong preference for de se readings, but where the possibility of a de re reading can be revealed by means of an appropriate discourse context. The findings from Pearson (2017) discussed above provide evidence of this for *he himself*, and (15c) suggests that ECM reflexives belong in this category too. I shall return to this issue in section 6. Note, however that the test is likely only applicable for overt pronouns/anaphors: if the expression in question is unpronounced, it cannot receive focal stress. With that in mind, in the next sub-section I introduce a further strategy for overcoming the de se bias, which is applicable to overt and covert expressions alike.

## 4.2 Motivating mistaken identity

A second type of context where the relevance of the de se reading is demoted is found in Pearson and Roeper (2020, 2022):

- (18) *Baking competition (want)* Mary, Bob and John all like baking. Their teacher, Mr. Smith, wanted to know which of them baked the best cookies. He had an idea. Each of them would bake a batch of cookies, and then one of them would taste them and decide which one he or she thought was the best. Mr. Smith would give a prize to whoever’s cookie was judged the best. Mary entered the contest, but since the prize was a trip to The Cheesecake Factory – a restaurant she dislikes – she had no desire to win. Mr. Smith picked Mary to be the judge, and decided that she should wear a blindfold for the tasting so she wouldn’t know whose cookie was whose. When Mary tasted her own cookie, she couldn’t tell that it was hers. She said, ‘This is the best cookie. I want whoever baked this cookie to win the prize.’
- a. Mary doesn’t realize it, but she wants to win the prize.
  - b. Mary doesn’t realize it, but she wants herself to win the prize.

Several features of the setup are noteworthy here. First, a motivation for the mistaken identity is provided: to ensure that the contest is fair, the judge is blindfolded. Consequently, the mistaken identity aspect of the scenario is not redundant even if a de re reading is assigned to the target sentence. If the view set out at the beginning of section 4 about the source of the apparent preference for de se readings is correct, then this feature of the setup should help to overcome this preference.

Secondly, the target sentences are addressed to an interlocuter who knows about the use of the blindfold and is therefore likely to assume that the speaker did not know that she was talking about

herself. The addressee is therefore likely to know that if those sentences are true, they are true only on a *de re* reading. Thirdly, the question under discussion in this dialogue is ‘Who should win the prize?’ Given the rules of the game, it follows from what Mary said that she should win the prize; whether she knew that she was talking about herself is irrelevant to this question. Finally, the sentences are preceded by what Pearson and Roeper call ‘ignorance disclaimers’, which signal that there is some aspect of Mary’s mental state that she is unaware of – in this case, she has a desire about herself (that *she* wins the prize) but is unaware that her desire is directed towards herself.

One might expect that these contextual factors would be successful in overcoming the presumption of relevance for the mistaken identity feature that I have proposed underlies the tendency to interpret *de se/de re* ambiguous reports on their *de se* reading. If so, then a majority of speakers should judge (18b) true. At the same time, since the grammar determines that PRO is unambiguously *de se*, the sentence should be impervious to contextual manipulations. Sure enough, (18a) was uniformly judged false, whereas (18b) was judged true by four out of six speakers.<sup>19</sup>

This is not quite the whole story though. Pearson and Roeper found very different results when the embedding predicate was *claim* rather than *want*.

(19) *Baking competition (claim)*

Mary, Bob and John all like baking. Their teacher, Mr. Smith, wanted to know which of them baked the best cookies. He had an idea. Each of them would bake a batch of cookies, and then one of them would taste them and decide which one he or she thought was the best. Mr. Smith would give a prize to whoever’s cookie was judged the best. Mr. Smith picked Mary to be the judge, and decided that she should wear a blindfold for the tasting so she wouldn’t know whose cookie was whose. When Mary tasted her own cookie, she couldn’t tell that it was hers. She said, ‘This is the best cookie. Whoever baked this cookie deserves the prize.’ Sam had been keeping track of who made which cookie, so that he could go and tell the results to Mr. Smith. He went to Mr. Smith and said:

a. ‘Mary didn’t realize it, but she claimed to deserve the prize.’

b. ‘Mary didn’t realize it, but she<sub>i</sub> claimed that she<sub>i</sub> deserves the prize.’

[Pearson and Roeper 2022]

Surprisingly, 6 out of 6 native English speakers judged (19a) to be true. Pearson and Roeper (2022) argue that this is not evidence that PRO has a *de re* reading; if it did, then it would be mysterious why (18a) is judged false. Rather, under appropriate contextual pressure, judgments may mimic *de re* readings, due to a kind of pragmatic loosening of the interpretation which Pearson and Roeper refer to as an ‘excluded entailment’. They argue that the interpretive process at work here is the pragmatic analogue of the loosening associated with modifiers such as *in effect*, which can also generate what they call *apparent de re* readings of control sentences.<sup>20</sup> For example, (20) was judged true in *Baking competition* by 7 out of 11 speakers asked.

(20) In effect, Mary claimed to deserve the prize.

<sup>19</sup>I take this as evidence against Moulton’s (2008) claim that ECM reflexives do not allow a *de re* reading under *want*; I assume instead that ECM reflexives uniformly allow such readings, given a suitable discourse context.

<sup>20</sup>As far as I know this was first observed by Wataru Uegaki in a social media post.

Yet crucially, Pearson and Roeper found that when (18a) is modified by *in effect* only a minority (2 out of 6) of speakers judge the sentence true:

(21) In effect, Mary wants to win the prize.

Pearson and Roeper take it that the pair of contrasts between (18a) and (19a) on the one hand and (20) and (21) on the other is evidence that where we find apparent de re readings with PRO, they are in fact cases of *in effect* uses.<sup>21</sup> Pearson and Roeper argue that the crucial difference between *want* and *claim* lies in the fact that *want*-sentences with control infinitives are what they call ‘doubly de se’, an observation with roots in the following example due to Lakoff:

(22) I wanted to be President, but I didn’t want myself to be President. [Lakoff 1972]

Lakoff observed that (22) is not contradictory: it can be true in a context where the speaker had the selfish ambition of becoming President but didn’t think that it would be good for her country if she was President. Pearson and Roeper propose that it is this that prevents *in effect* modification or *in effect* uses from giving rise to apparent de re readings with obligatory control sentences involving *want*. On their analysis, ‘In effect p’ (or an *in effect* use of a sentence p) communicates roughly that the relevant outcomes of what took place are equivalent to the relevant outcomes of p, had p taken place. (20) is true, for instance, not because Mary claimed to deserve the prize – she didn’t – but because she *may as well have done*: either way, Mary gets the prize, given the rules of the game. By contrast, the relevant outcomes of Mary wanting the baker of the cookie she liked best to win the prize are not the same as the relevant outcomes of her wanting to win the prize: as Lakoff’s example teaches us, the latter has no bearing on whether Mary ought to win the prize.

Based on these observations, I propose a strategy of overcoming the de se bias by (i) setting up contexts such as that employed by Pearson and Roeper, where the relevance of the de se/de re distinction is reduced and the target sentence is preceded by an ignorance disclaimer and (ii) employing an embedding predicate like *want* which seems to resist the apparent de re readings of PRO found by Pearson and Roeper with *claim*. If a desire report containing  $\alpha$  is found to be judged true in a context that has been set up following this strategy, then one should conclude that  $\alpha$  has a de re reading. Further research is of course needed in order to understand the source of ‘excluded entailments’ better, and to learn more about which predicates pattern with *want* in resisting this effect.<sup>22</sup> This can be diagnosed by modification with *in effect*: if the resulting sentence lacks an apparent de re reading, then the predicate in question can be classed with *want*, and therefore be pressed into service for this strategy.

<sup>21</sup>Jérémy Pasquereau (p.c.) notes that he has similar intuitions for *basically*; this modifier was not considered in Pearson and Roeper (2020, 2022), but I agree with his intuitions.

<sup>22</sup>This raises a question about Ewe and Amharic. Could the putative evidence for de re readings of logophors and shifty indexicals in these languages in fact be instances of pragmatic loosening? If the reasoning outlined in this sub-section is correct, this could be checked by constructing target sentences with *want*. Unfortunately such data has not been reported for either language, as far as I am aware. On the other hand, the data that we do have does not involve the kinds of contextual manipulations employed by Pearson and Roeper. Additionally, the availability of the strict reading for the Ewe logophor provides further evidence against the hypothesis that it is obligatorily read de se (see section 5).

## 5 A note on strict readings

A crucial fact about the logophoric pronoun in Ewe that I overlooked in Pearson (2013, 2015) is that it can receive strict readings when the attitude holder is a quantified DP containing *only*. As far as I know this was first noted in Culy (1994), who provides the following example.<sup>23</sup>

- (23) Kofi ko e-hose be Ama lɔ̃ yè  
 Kofi only 3SG-believe COMP Ama loves LOG  
 ‘Only Kofi believes that Ama loves him.’ [Culy 1994: 1082, ex (41a)]

Culy notes that (23) is ambiguous between a strict and a sloppy reading. On the strict reading, the sentence reports that Kofi is the only person who believes that Ama loves Kofi. On the sloppy reading, it means roughly that the property of being an *x* such that *x* believes that Ama loves *x* applies only to Kofi.

I was not aware of this data at the time of writing Pearson (2013, 2015). (Indeed the theory proposed there predicts that only the sloppy reading is available for *yè*, since the logophor is treated as an obligatorily bound variable. I believe that this is a fatal problem for that analysis.)

Bimpeh and Sode (2021) have also discussed the availability of the strict reading for *yè*, and the problems that it raises for previous analyses:<sup>24</sup>

- (24) Kofi<sub>i</sub> ko yé xɔ̃se be yè<sub>i</sub> kpɔ̃ ɲɔ̃li  
 Kofi only FOC believethat LOG see ghost  
 ‘Only Kofi believes that he saw a ghost.’  
 a. Strict reading: Kofi is the only person who believes that Kofi saw a ghost.  
 b. Sloppy reading: The property of being an *x* such that *x* believes that *x* saw a ghost applies only to Kofi. [Bimpeh and Sode 2021, ex (3)]

Data such as these are the crucial clue that we all missed that the logophoric pronoun in Ewe has a *de re* reading. I know of no theory that would predict that a logophor could (i) be unambiguously read *de se* and (ii) allow strict readings. On standard assumptions, the strict reading of (24) arises via the following LF.

- (25) [Only Kofi<sub>7</sub>] λx<sub>1</sub> [t<sub>1</sub> believes that yè<sub>7</sub> saw a ghost]

Crucially, *yè* is not coindexed with the abstractor that binds the subject trace - that would yield the sloppy reading - but instead is co-indexed with Kofi. As far as I am aware there is no theory that would assign *de se* truth conditions to this LF, to the exclusion of a *de re* reading.

In general then, I propose that if  $\alpha$  can receive a strict reading in the scope of an *only*-DP (or indeed with verb phrase ellipsis), this should be treated as evidence that  $\alpha$  allows a *de re* reading. If a strict reading is available for  $\alpha$  and a consultant judges an attitude report containing  $\alpha$  to be false

<sup>23</sup>I have reproduced this data following the orthographical conventions employed in Culy (1994). These are different from those in Pearson (2013, 2015).

<sup>24</sup>Satik (2020) also reports that *yè* can receive a strict reading under VP ellipsis in Anlo Ewe. I focus on the case of *only* for expository purposes.



in a mistaken identity scenario, this should be taken to reflect a bias towards de se readings rather than to demonstrate that the de re reading is out altogether.<sup>25</sup>

Note that the reverse does not hold: if  $\alpha$  only has a sloppy reading it should not be concluded that it is unambiguously de se. To illustrate this, consider the LF for the sloppy reading of (24).

(26) [Only Kofi<sub>7</sub>]  $\lambda x_1$  [t<sub>1</sub> believes that yè<sub>1</sub> saw a ghost]

Here, yè is coindexed with the abstractor that binds the subject trace, yielding the property of being an  $x$  such that  $x$  believes that  $x$  saw a ghost. On standard assumptions, this configuration, where the binder is higher than the attitude verb, generates the de re reading. In principle then there could be an element  $\alpha$  that cannot receive a strict reading, but can be read de re. So if one finds that  $\alpha$  only has a sloppy reading, one cannot conclude anything about whether or not it is obligatorily de se.

## 6 A possible alternative: establishing a baseline

I have proposed that the contexts typically used to test for de se/de re ambiguities tend to bias the consultant towards interpreting the target sentence on the de se reading, owing to Modified Chekov's Gun, which in turn gives rise to a presumption of relevance for the mistaken identity feature. I have also described two ways of modifying these contexts that can help to overcome this difficulty. In this section, I consider an alternative possible strategy for addressing the de se bias, based on the use of minimal pairs to establish a baseline against which to compare judgments for the target sentence. Since many claims in the literature regarding obligatory de se interpretation of some pronoun/anaphor are grounded in data involving such minimal pairs, we can ask whether this strategy is sufficient to show that an anaphoric element is obligatorily de se, despite the de se bias. In this section, I will argue that it is not.

Elementary methodological considerations tell us that linguistic intuitions for a given target sentence  $S$  should be compared against those for a minimally different baseline sentence  $S'$ , in order to establish that conclusions drawn on the basis of intuitions about  $S$  are warranted.<sup>26</sup> Here then is a tempting line of reasoning. Suppose that we have an attitude report  $S$  that contains an embedded pronoun/anaphor  $\alpha$  that corresponds to the attitude holder. Let  $S'$  be an attitude report that is just like  $S$  except that  $\alpha$  is replaced with another appropriate pronoun/anaphor  $\alpha'$  (*mutatis mutandis*). We elicit judgments concerning  $S$  and  $S'$  in a mistaken identity scenario and find that the speaker judges  $S$  false and  $S'$  true. It might seem that we should then be justified in concluding that  $\alpha$  is unambiguously de se, despite the worries about de se bias raised in this paper.

To illustrate, let us return to the Ian McKellen example introduced earlier. I'll adapt the scenario slightly as follows, in order to make use of the control predicate *claim*.

(27) Scenario: In 2007, the actor Ian McKellen played the role of King Lear – a role for which he was required to grow a beard. On being complimented on the beard in a television interview, McKellen responded, 'It's taken a year to grow, I hate it. I walk along the street and see this geezer in shop windows and I say, "Oh dear, he looks terrible!"'

<sup>25</sup>Of course, if someone were to come up with a compelling story about how an expression could be obligatorily de se and still allow strict readings, then this proposal would have to be re-thought.

<sup>26</sup>For relevant discussion pertinent to semantic data in particular, see Tonhauser and Matthewson (2016).

- a. McKellen claimed PRO to look terrible.
- b. McKellen claimed that he looked terrible.

Many native English speakers are willing to grant that a sentence like (27b) is true in such a scenario but insist that (27a) is false. This might seem on first sight like robust evidence that PRO is unambiguously read *de se*. Considered on its own, the judgment for (27a) is compatible with the sentence simply being interpreted with respect to the *de se* reading, rather than showing that it has no *de re* reading. But if this were so, then we would expect the *de se* bias to carry over to (27b) too, leading to it being judged false. One might conclude then that the willingness of a speaker to judge (27b) but not (27a) to be true is sufficient evidence that PRO is unambiguously *de se*, despite the *de se* bias.

The problem with this argumentation can be seen by replacing the control infinitive in (27a) with a finite clause containing *he himself*.

(28) McKellen claimed that he himself looked terrible.

Many speakers (myself included) detect a contrast between (27b) and (28): (27b) is true (or can be), whereas (28) seems false. So we have compared a report containing *he himself* against a minimally different variant containing the ordinary pronoun *he*, identified a contrast with respect to a mistaken identity scenario, and seem justified in concluding that *he himself*, like PRO, is unambiguously read *de se*. But as we saw in section 4.1, this is incorrect: the availability of a *de re* reading for *he himself* is revealed once an appropriate discourse context is set up. So while minimal pairs are an important tool, they are not on their own sufficient to guarantee that the *de se* bias is overcome.

Example (27) illustrated a strategy of comparing the target sentence with a baseline sentence that we have reason to believe has a *de re* reading. A related strategy is to compare the sentence with a baseline sentence that we have reason to believe is unambiguously *de se*. Suppose for example, that we were interested in finding out whether the ordinary English pronoun *he* is unambiguously *de se*. When presented in isolation, (27b) might very well be judged by a speaker to be false. Again, this would be compatible with the sentence being ambiguous between a *de se* and a *de re* reading. If instead we present the pair of sentences (27a) and (27b), the speaker might be more inclined to judge (b) true; the juxtaposition with a sentence that is clearly unambiguously *de se* serves to highlight a reading that might otherwise have gone undetected.<sup>27</sup> On this basis, it might seem that we could then conclude that the hypothesis that *he* is unambiguously *de se* is false, despite appearances to the contrary when a target sentence containing this pronoun is presented in isolation.

Here again, the limitations of this strategy are revealed by considering the case of *he himself*. When (28) is presented in isolation, one is inclined to judge it false in the scenario under consideration. But this judgment persists even when (28) is juxtaposed with its counterpart with PRO, despite the fact that *he himself* does have a *de re* reading.

None of this is to deny that when testing for *de se/de re* ambiguities it is good practice to set up a baseline condition for comparison with the target sentence. But it does suggest that a contrast of the type exemplified in (27) is not in and of itself sufficient evidence that the pronoun/anaphor of interest lacks a *de re* reading. After all, a contrast of this type is compatible with the availability of a

<sup>27</sup>Idan Landau (p.c.) remarked at the Workshop on Logophoricity and Perspectivization in Wackershofen in 2018 that in his teaching he has found this strategy effective in helping students to detect the contrast between ordinary pronouns and PRO.

de re reading for *both* the pronoun/anaphor of interest  $\alpha$  *and* the expression  $\alpha'$  that it is compared to, albeit with  $\alpha$  being associated with a *stronger* de se preference than  $\alpha'$ . This seems to be the case for *he himself* and for ECM reflexives: the de re reading can be brought out by setting up an appropriate discourse context that renders the use of these elements felicitous even when the awareness condition associated with de se is not met. I conclude that the detection of a contrast when comparing the target sentence with an appropriate baseline is not in itself sufficient to address the problem of de se bias.

## 7 Conclusion

In this paper I have been concerned with the predicament of the fieldworker who wants to test the hypothesis that some pronoun or anaphor  $\alpha$  is unambiguously interpreted de se. As a case study, I first discussed the claim that the logophoric pronoun in Ewe is obligatorily de se – a notion that until recently had gained considerable traction in the literature. I then described how in my own consultant work I demonstrated that this hypothesis is false. The strategy involved presenting a mistaken identity scenario and asking whether the target sentence containing a logophor is true or false. A judgement of truth is considered evidence that the logophor has a de re reading (providing that one is confident that the consultant has understood the scenario and the task correctly). But a ‘false’ response doesn’t necessarily show that the logophor is unambiguously read de se, since (i) such a response is compatible with the logophor being ambiguous between a de se and a de re reading, (ii) the contexts typically used to test for de se/de re ambiguity tend to generate a bias in favour of the de se reading and (iii) some anaphors, such as ECM reflexives and *he himself*, trigger a particularly strong preference for the de se reading in many contexts.

The lesson is that testing for de se and de re readings in an unfamiliar language is a difficult task, but not for the reasons that are often emphasised. It is true that the scenarios involved are often strange and challenging to reason or introspect about. But with careful collaboration with the consultant, this problem can be overcome: what we call ‘scenarios’ or ‘contexts’ are after all merely stories, and engaging with stories is something we all do all the time. The greater problem is that of demonstrating the *absence* of a reading – a problem rendered more acute by the fact that the reading of interest may be dispreferred in precisely the type of scenario that is needed to test for de se/de re ambiguity.

When a speaker gives a judgment for a sentence S that is unexpected if S is assigned reading B, the possibility arises that S may be ambiguous between B and some other reading A, and that the judgment given demonstrates that the speaker has interpreted S with respect to reading A, not that S lacks reading B altogether. As the case of de se illustrates, this problem is particularly serious in cases where there is a bias towards reading A, which may mask the availability of reading B.

In this paper, I have identified one possible source of such a bias: the presumption that a particular feature of a scenario is relevant. If interpreting S with respect to A but not B enables this presumption to be preserved, then this may give rise to a bias in favour of A. I have argued that this issue arises when testing for de se/de re ambiguities, unless one finds a way to either (i) embed the target sentence in a discourse where the mistaken identity feature of the scenario is rendered irrelevant to the current question (thereby defeating the presumption of relevance) or (ii) build an independent motivation for the mistaken identity into the narrative of the scenario (thereby ensuring that the presumption of relevance is maintained even if the target sentence is read de re). A goal for future work should be to examine other corners of the grammar to see what other ambiguities may be masked by features of the presented scenarios in an analogous way to what I have described here.

Much insightful and influential work on the semantics of de se reports is predicated on claims about which expressions are unambiguously de se. These assertions presuppose the existence of a robust procedure for establishing that a given expression lacks a de re reading, and yet to my knowledge, no such methodology has ever been explicitly proposed. One criterion for determining the adequacy of such a methodology is that it should be able to establish whether the data is indicative of the unavailability of the de re reading, or merely reflects bias towards de se readings in the contexts presented. I hope that the strategies described in this paper can serve as a step towards developing a methodology that meets this criterion.

## References

- Anand, Pranav. 2006. De De Se. Doctoral dissertation, Massachusetts Institute of Technology. Retrieved from <http://hdl.handle.net/1721.1/37418>.
- Bimpeh, Abigail Anne. 2019. Default de se: The interpretation of the Ewe logophor. In *Proceedings of Triple A5*, ed. M. Ryan Bochnak, Miriam Butt, Erlinde Meertens and Mark-Matthias Zymla, 1-15. Universität Tübingen library publication system.
- Bimpeh, Abigail Anne. 2021. Logophoricity in Ewe: An Empirical-Semantic Assessment of yè. Doctoral dissertation, Johann Wolfgang Goethe-Universität zu Frankfurt am Main.
- Bimpeh, Abigail Anne and Frank Sode. 2021. Evidence against de se binding: Strict readings of the logophoric pronoun in Ewe. In *Proceedings of TripleA6*, ed. Neil Banerjee and Verena Hehl, 1-16. MIT Working Papers in Linguistics.
- Chierchia, Gennaro. 1990. Anaphora and attitudes de se. In *Semantics and contextual expression*, ed. Renate Bartsch, Johan van Benthem and Peter van Emde Boas, 1–32. Dordrecht: Foris.
- Culy, Christopher. 1994. Aspects of logophoric marking. *Linguistics* 32. 1055–1094.
- Deal, Amy Rose. 2020. *A Theory of Indexical Shift: Meaning, Grammar, and Crosslinguistic Variation*. Cambridge, MA: MIT Press.
- Geach, Peter T. 1957. On beliefs about oneself. *Analysis* 18, 23-24.
- Grice, H. Paul. 1970. Logic and conversation. Reprinted in Grice, H. Paul. (1989). *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.
- Groenendijk, Jeroen, and Martin Stokhof. 1984. Studies on the Semantics of Questions and the Pragmatics of Answers. Doctoral dissertation, University of Amsterdam.
- Haida, Andreas. 2009. (Proto-)logophoricity in Tangale. Handout.
- Hamblin, Charles L. 1958. Questions. *Australasian Journal of Philosophy*, 36: 159–168.
- Heim, Irene. 2002. Features of pronouns in semantics and morphology. Handout of a talk given at USC.
- Hintikka, Jaakko. 1969. Semantics for propositional attitudes. In *Models for Modalities*, 87–111. Springer. Retrieved from [https://doi.org/10.1007/978-94-010-1711-4\\_6](https://doi.org/10.1007/978-94-010-1711-4_6).
- Hornstein, Norbert and Paul Pietroski. 2010. Obligatory control and local reflexives: copies as vehicles for de se readings. In *Movement Theory of Control*, ed. Hornstein, Norbert and Maria Polinsky. John Benjamins Publishing.
- Huang, C.-T. James, and C.-S. Luther Liu. 2001. Logophoricity, Attitudes and ziji at the Interface. In *Syntax and semantics: Long distance reflexives*, ed. Peter Cole, Gabriella Hermon, and C.-T. James Huang, 150–195. Academic Press.
- Kusumoto, Kiyomi. 1998. Tenses as Logophoric Pronouns. Hand-out of talk given at the MIT/ UConn/UMass Semantics Workshop, October 31, 1998 (UConn).

- Lakoff, George. 1970. Linguistics and natural logic. *Synthese* 22(1). 151–271. Retrieved from <https://doi.org/10.1007/BF00413602>.
- Landau, Idan. 2015. A two-tiered theory of control, vol. 71. MIT Press. Retrieved from <https://doi.org/10.7551/mitpress/9780262028851.001.0001>.
- Landau, Idan. 2018. Direct variable binding and agreement in obligatory control. In Pritty Patel-Grosz, Patrick Georg Grosz and Sarah Zobel (eds.), *Pronouns in embedded contexts at the syntax-semantics interface*, 1–41. Cham: Springer International Publishing. Retrieved from [https://doi.org/10.1007/978-3-319-56706-8\\_1](https://doi.org/10.1007/978-3-319-56706-8_1).
- Maier, Emar. 2006. Belief in context: Towards a unified semantics of de re and de se attitude reports. Doctoral dissertation, Radboud Universiteit Nijmegen. Retrieved from <https://sites.google.com/site/emarmaier/publications>.
- Maier, Emar. 2009. Presupposing acquaintance: a unified semantics for de dicto, de re and de se belief reports. *Linguistics and Philosophy* 32(5). 429–474. Retrieved from <https://doi.org/10.1007/s10988-010-9065-2>.
- Maier, Emar. 2011. On the roads to de se. In *Semantics and Linguistic Theory*, vol. 21, 393–412. Retrieved from <https://journals.linguisticsociety.org/proceedings/index.php/SALT/article/viewFile/2611/2358>.
- Malamud, Sophia. 2006. Semantics and Pragmatics of Arbitrariness. Doctoral dissertation, University of Pennsylvania.
- Morgan, Jerry. 1970. On the criterion of identity for noun phrase deletion. In *Sixth Regional Meeting, Chicago Linguistic Society*, 380–389.
- Moulton, Keir. 2008. Strong reflexivity and attitudes de se: Evidence from ECM. Handout, CONSOLE XIV E.H.U.-U of the Basque Country.
- O’Neill, Teresa. 2016. The distribution of the Danyi Ewe logophor *yi*. Talk given at the *Annual Meeting of the Linguistic Society of America*. Retrieved from [https://static1.squarespace.com/static/55e3a4d0e4b06cce652896c2/t/56a2ba800e4c11785794d0bd/1453505153503/LSA\\_O%27Neill\\_slides.pdf](https://static1.squarespace.com/static/55e3a4d0e4b06cce652896c2/t/56a2ba800e4c11785794d0bd/1453505153503/LSA_O%27Neill_slides.pdf).
- Oshima, David Y. 2004. Zibun revisited: empathy, logophoricity, and binding. In *University of Washington Working Papers in Linguistics*, volume 23, 175–190.
- Park, Yangsook. 2018. Attitudes de se and logophoricity. Doctoral dissertation, University of Massachusetts Amherst.
- Patel-Grosz, Pritty. 2020. Pronominal typology and the de se/de re distinction. *Linguistics and Philosophy* 43: 537–587. Retrieved from [https://doi.org/10.1007/s10988-019-09274-7\(0123456789\(\).,-vo](https://doi.org/10.1007/s10988-019-09274-7(0123456789().,-vo)
- Pearson, Hazel. 2013. The sense of self: Topics in the semantics of de se expressions. Doctoral dissertation, Harvard University. Retrieved from <http://semanticsarchive.net/Archive/WNhNjVjO/>.
- Pearson, Hazel. 2015. The interpretation of the logophoric pronoun in Ewe. *Natural Language Semantics* 23(2). 77–118. Retrieved from <https://doi.org/10.1007/s11050-015-9112-1>.
- Pearson, Hazel. 2017. He himself and I. *Snippets* 31. 20–21. Retrieved from <http://dx.doi.org/10.7358/snip-2017-031-pear>
- Pearson, Hazel. 2018. Counterfactual de se. *Semantics and Pragmatics* 11(2). Retrieved from <https://doi.org/10.3765/sp.11.2>.
- Pearson, Hazel and Tom Roeper. 2020. Three puzzles about de se. In *Making Worlds Accessible: Festschrift for Angelika Kratzer*, ed. Frana, Ilaria, Menendez-Benito, Paula and Rajesh

- Bhatt, 193-216. UMass ScholarWorks. Retrieved from <https://doi.org/10.7275/w10a-pv24>
- Pearson, Hazel and Tom Roeper. 2022. Excluded entailments and the de se/de re partition. *Inquiry: An Interdisciplinary Journal of Philosophy*. 1-29.
- Percus, Orin and Uli Sauerland. 2003. Pronoun movement in dream reports. In *Proceedings of NELS 33*, ed. Kadowaki, Makoto and Shigeto Kawahara, 265–284. GLSA.
- Podobryaev, Alexander. 2014. Persons, Imposters, and Monsters. Doctoral dissertation, Massachusetts Institute of Technology.
- Reinhart, Tanya. 1990. Self-representation. Unpublished manuscript. Retrieved from [http://www.let.uu.nl/~tanya.reinhart/personal/Papers/pdf/De\\_se\\_91.wp.pdf](http://www.let.uu.nl/~tanya.reinhart/personal/Papers/pdf/De_se_91.wp.pdf).
- Roberts, Craige. 2012. Information structure in discourse: towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5. 1-69.
- Santorio, Paolo. 2014. On the plurality of indices. Manuscript, University of Leeds. Retrieved from <http://paolosantorio.net/opi.draft3b.pdf>.
- Satik, Deniz. 2019. The interpretation of the nonfinite logophoric pronoun in Anlo Ewe. Slides of talk given at *TripleA*.
- Satik, Deniz. 2020. Logophoric pronouns are not (always) inherently logophoric. In *Proceedings of the Fifty-fifth Annual Meeting of the Chicago Linguistic Society*, ed. Ömer Eren, Asimina Giannoula, Sam Gray, Chi-Dat Lam (Daniel), Aurora Martinez Del Rio, 381-396. Chicago, IL.
- Schlenker, Philippe. 1999. Propositional attitudes and indexicality: A cross categorial approach. Doctoral dissertation, Massachusetts Institute of Technology. Retrieved from <http://hdl.handle.net/1721.1/9353>.
- Schlenker, Philippe. 2000. Propositional attitudes and indexicality: A cross categorial approach. Doctoral dissertation, Massachusetts Institute of Technology. [Revised version].
- Schlenker, Philippe. 2003. A plea for monsters. *Linguistics and Philosophy* 26(1). 29–120. Retrieved from <https://doi.org/10.1023/A:1022225203544>.
- Schlenker, Philippe. 2011. Indexicality and de se reports. In *Semantics: an international handbook of contemporary research*, ed. Klaus von Stechow, Claudia Maienborn and Paul Portner, 1561–1603. Berlin: de Gruyter.
- Schlenker, Philippe. 2017. Super monsters I: Attitude and Action Role Shift in sign language. *Semantics and Pragmatics* 10(9). [Early access] Retrieved from <http://dx.doi.org/10.3765/sp.10.9>.
- von Stechow, Armin. 2003. Feature deletion under semantic binding: Tense, person, and mood under verbal quantifiers. In *Proceedings of NELS 33*, ed. Kadowaki, Makoto and Shigeto Kawahara, 379–404.
- Sudo, Yasutada. 2016. Person Indexicals in Uyghur Indexical Shifting. In *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 36, ed. Nicholas Rolle, Jeremy Steffman and John Sylak-Glassman, 441-456. Berkeley, CA: Berkeley Linguistics Society.
- Tonhauser, Judith and Lisa Matthewson. 2016. Empirical evidence in research on meaning. Retrieved from [lingbuzz/002595](http://lingbuzz/002595)
- Zu, Vera. 2018. Discourse Participants and the Structural Representation of the Context. Doctoral dissertation, New York University.