Easy as ABC. Functional-pragmatic factors explain “binding-principle” constraints on pronoun interpretation: Evidence from nine pre-registered rating studies.

DOI:
10.2139/ssrn.5014467

Easy as ABC. Functional-pragmatic factors explain "binding-principle" constraints on pronoun interpretation: Evidence from nine pre-registered rating studies.

1. Introduction

Pronouns are having their moment. Whether it's *he/him*, *she/her* or *they/them*, these little words (in a development that no one could have predicted as recently as 10 years ago) now lie at the very heart of the culture wars (e.g., McWhorter, 2024; Tucker & Jones, 2023). But a very different battle around pronouns has, for almost half a century, stood at the heart of what historians of science have referred to – only half-jokingly – as the "linguistics wars" (Harris, 1993). This debate over pronouns may at first seem arcane; but it is a crucial test case for competing explanations of our knowledge of *language* – a strong candidate for both our species' greatest achievement and its defining characteristic. Does our knowledge of language consist of grammatical rules that are too abstract to be learned, and must therefore be hardwired from birth; or does it emerge from our attempts to understand what our fellow speakers are trying to communicate? Just as in the culture wars, the debate all comes down to pronouns.

How so? Well, consider the following sentences (which, together, could form a complete short story):

(a) Samuel kicked himself.
(b) Samuel kicked him.
(c) He kicked Samuel.

For all fluent English speakers – including, in the main, older children (e.g., Bergmann et al., 2011; Chien & Wexler, 1990; Clackson et al., 2011; Kujiper et al., 2021; Matthews et al., 2009; McKee, 1992) – it is beyond obvious that, in (a) *himself* can only mean *Samuel*, while *him* in (b) and *he* in (c) cannot mean *Samuel*. But where do these "rules" come from?

Since at least Chomsky (1981)[1], one popular view (Adger & Svenonius, 2015; Crain et al., 2017; Reinhart, 1983; Reinhart & Reuland, 1993; Reuland, 2011; Trueswell, 2014), which we will call the ***formalist*** view, has been that these rules are part of speakers' highly abstract knowledge of the grammatical rules of their language (sometimes called "syntax"). For example, simplifying to avoid complex linguistic terminology, one rule states that – for SUBJECT VERB OBJECT – sentences like (a) *Samuel kicked himself*, a reflexive pronoun (*himself, herself, themself* etc.) in OBJECT position **must** take its meaning from the

SUBJECT of the sentence (hence, *himself* must mean *Samuel*). A second rule states that, for sentences of this type (e.g., [b] *Samuel kicked him*), a nonreflexive pronoun in OBJECT position (e.g., *him, her, me*) **must not** take its meaning from the SUBJECT of the sentence (hence, *him* cannot mean *Samuel*). A third rule states that, for sentences of this type (e.g., [c] *He kicked Samuel*) a full-noun-phrase (i.e., something that is NOT a pronoun) in OBJECT position (e.g., *Samuel, the man*) **must not** take its meaning from whatever is in the SUBJECT position (hence, *he* cannot mean *Samuel*).

The crucial thing to note about these formalist rules (traditionally known as *Binding Principles A, B* and *C*) is that they are formulated solely in terms of abstract grammatical categories like *SUBJECT, OBJECT, reflexive pronoun* and *nonreflexive pronoun*. Indeed, the dominant formalist view has long been that these rules are *so* abstract – so highly removed from the actual sentences people say and hear – that they cannot be learned. Rather "these binding principles, and other linguistic constraints, are seen to be part of the innately specified Universal Grammar" (Crain et al., 2017., p127), with which all humans are born.

The opposing ***functionalist*** view (e.g., Ambridge & Lieven, 2011; Ambridge et al., 2014; Bickerton, 1975; Bolinger, 1979; Cole et al., 2015; Jackendoff, 1992; Kuno, 1987; Lakoff, 1968; Levinson, 1987; MacWhinney, 2008, 2009; Matthews et al., 2009; van Hoek, 1995; Van Valin & LaPolla, 1997) holds that this type of knowledge – that *himself* in (a) must mean *Samuel*, while *him* in (b) and *he* in (c) cannot mean *Samuel* – does not take the form of abstract grammatical rules. Rather, this type of knowledge comes from the listener's inference about *what the speaker meant to convey*, such that speakers tend to choose their words carefully, taking the listener's knowledge into account (e.g., Grice, 1975).

For example, under this functionalist account, the reasoning that underlies the interpretation of a ("Principle A") sentence such as *Samuel kicked himself* is something like the following[2]: "In the normal state of affairs, the person who does something (e.g., kicking) and the person who has something done to them (e.g., being kicked) are different people. However, instead of using the normal all-purpose word for a male who has something done to them – *him* – the speaker has gone out of their way to use a much more unusual word - *himself*. This word (a) is longer and more complex, (b) is much less frequent (in the Corpus of Contemporary American English, 88% of uses of *him* or *himself* are *him*) and (c) has a much more specialized meaning: a male who instigates the event of which he is also the direct target (and/or who is seen from his own point of view[3]). Thus, the speaker clearly intends me to infer that, unusually, the kicker and the person being kicked are one and the

same". Of course, listeners do not need to follow this complete chain of reasoning each time they hear such a sentence; rather, in the course of hearing many such sentences, they will have developed proceduralized interpretation pathways. Armed with this knowledge, a listener who hears a ("Principle B") sentence such *Samuel kicked him* need reason only "If the speaker meant 'himself' [as defined above] they would have said *himself*."

Finally, the reasoning that underlies the interpretation of a ("Principle C") sentence such as *He kicked Samuel* is something like the following "As the starting point, the speaker established the topic of the sentence – that is, who we're talking about – using *He* (Arnold et al., 2013; Kuno, 1987). And I know that speakers avoid using full names when – as is the case here – doing so would be more informative than necessary given what listeners already know (i.e., the repeated name penalty; Gordon et al., 1993). This means they're confident we both already know who we're talking about (e.g., *Yusuf*). But they certainly can't mean *Samuel*, because in this counterfactual world where we'd both established that we *were* talking about *Samuel* – which is what allows the speaker to say *he* instead of *Samuel* in the first place – it would be totally weird to then specify *Samuel* in the same sentence. And in any case, if they did mean *Samuel* they would have just said *Samuel kicked himself* (or, if we already both knew we were talking about *Samuel*, *He kicked himself*)". Again, this chain of reasoning has likely been proceduralized into a quasi-Gricean principle of reference (e.g., Gernsbacher, 1990; Klatzky et al., 2008; MacWhinney, 1977; Pecher & Zwaan, 2005), rather than needing to be computed in real time.[4]

Turning to experimental work, formalist studies tend to simply assume as a starting point that these formal rules – or binding principles – exist, and investigate (using mainly real-time online processing methods) how they interact with morphological cues like gender and number agreement (Nicol & Swinney, 1989; Clifton et al., 1999; Dillon et al., 2013; Chow et al., 2014; Cunnings and Sturt, 2014; also see Parker & Phillips, 2017; Jäger et al., 2020), or when they are mastered by children (e.g., Bergmann et al, 2011; Chien & Wexler, 1990; Clackson et al., 2011; Kujiper et al., 2021; McKee, 1992). Papers in the functionalist camp have so far rarely reported empirical studies at all, and instead mainly argue from examples; providing discovered or invented sentences that accord with functional principles and seem, potentially, to violate formal ones[5]. Given the pivotal theoretical status of "binding-principle" type sentence interpretations, it is surprising that neither camp has run experiments that investigate pronoun interpretation outside the scope of the "core" sentence

types exemplified by [a-c], for which formalist and functionalist accounts – despite their very different underlying assumptions and explanations – make identical predictions.

It is important to clarify from the outset that our goal in the present investigation is **not** to directly compare the predictions of formalist and functionalist accounts. One difficulty in doing so is that, for many sentence types (e.g., [a-c]) the two make identical predictions. Another difficulty is that formalist accounts have moved on from the "pure" syntactic approaches of the 1980s (e.g., Chomsky, 1981; 1986) and diversified by means of an array of "add-ons", many of which could be seen as incorporating in some form functionalist-style principles of reference (e.g., 'Rule 1' from Grodzinsky & Reinhart, 1993; 'Principle P' from Chien & Wexler, 1990; and the 'guise creation' accounts of Lidz et al., 2021; Reinhart, 1983; Thornton & Wexler, 1999; see Jackendoff, 1992).

Instead, then, our goal in the present studies is to investigate – setting aside relatively straightforward cases such as [a-c] – the extent to which a functional-pragmatic account alone can explain patterns of pronoun interpretation traditionally attributed to binding principles. To achieve this goal, we "control out" binding principles by selecting, within each of nine individual studies, a set of stimulus sentences for which the relevant binding principle (in its pure form) would make identical predictions, but for which functional-pragmatic accounts would predict widely different interpretations (i.e., ranging – sentence-by-sentence - from *Definitely NOT Samuel* to *Definitely Samuel*, depending on participants' inferences regarding a hypothetical speaker's intended meaning).

Specifically, on our interpretation, and assuming a "pure" syntactic version of the account, binding principles predict the following: For Experiments 1-3, Principle A predicts that for a sentence such as *Samuel asked/told Oliver about himself, himself* can be interpreted only as *Samuel* (modified formulations do allow *himself* to refer to either *Samuel* or *Oliver*, but make no predictions as to which[6]). For Experiments 4-6, Principle B predicts that for a sentence such as *Samuel asked Oliver about the picture of him*, *him* cannot refer to either *Samuel* or *Oliver*, and so must refer to a third person (again, modified formulations attempt to explain why *Samuel* and/or *Oliver* interpretations are possible for many speakers – see Jaeger et al (2004) for discussion – but make no predictions as to which will be preferred). For Experiments 7-9, Principle C predicts that for a sentence such as *He was exercising in the gym when Yusuf started whistling*, he cannot refer to *Yusuf* (here we are not aware of any proposed modifications that would render this interpretation possible[7]).

In contrast, taking Studies 1-3 (e.g., *Samuel asked/told Oliver about himself*) as an example of the functionalist account, applying the same single underlying construct of

plausibility as it would explain for sentence example [a][3] predicts that we will see a wide range of interpretations of *himself*– from *Definitely Samuel* to *Definitely Oliver*. Crucially, though, these interpretations are predicted not to be in free variation, but to relate to the meaning most likely intended by the (hypothetical) speaker. All other things being equal, *himself* is equally likely to refer to 'Samuel' or 'Oliver'. But in the real world all other things are never equal. For example, since people generally know more about themselves than others, *himself* is more likely to mean *Samuel* in *Samuel told Oliver about himself*, but *Oliver* in *Samuel asked Oliver about himself.* Conversely, if a speaker says *Someone told Oliver about himself*, the use of *Someone* (c.f., *Samuel*) suggests that the identity of the teller is unknown and or unimportant, and thus that *himself* probably refers to *Oliver*. Again, we are not suggesting that listeners compute these inferences in real time; these interpretation biases are likely to be baked into listeners' knowledge of the verbs *ask* and *tell*; just as implicit causality is baked into verbs like *criticize* and *amaze* (e.g., McDonald & MacWhinney, 1995)[8].

In this way, we aim to hold constant – and thus control out – the predictions of formalist (binding principle) accounts, in order to see just how much of the data can be explained by functional-pragmatic factors. It is important to be clear, of course, that no evidence in support of functionalist accounts could ever *disprove* the existence of formalist-style (possibly innate) rules or binding principles, in just the same way that no number of white swans could ever disprove the existence of black swans (also see Dąbrowska, 2015; MacWhinney, 2005; Xu, 2019). No empirical evidence could *ever in principle* disprove the claim that abstract formal principles are underlyingly present, with participants' sentence interpretations simply modified or influenced by the functionalist factors investigated. And, indeed, no formalist account would deny that functionalist discourse and pragmatic factors *are* important too (Newmeyer, 2010). What we *could* potentially show – and, we will argue, what the data do show – is that together, the functionalist factors that we investigate explain such a large proportion of variability in participants' judgments (e.g., from *Definitely Samuel* to *Definitely Oliver*) that, at least for these particular sentence types, there is little that would require recourse to other factors (potentially including formal rules and principles).

1.1. Ethics

All studies were approved by the University of Liverpool Research Ethics Committee (10230 - Pronoun Interpretation: Ratings on a 1-100 scale). All participants provided informed consent for their participation.

1.2. Preregistration and Data availability

All studies were preregistered on individual OSF pages (see https://osf.io/9stxh/? [Experiment 1], https://osf.io/9fxmh/ [Experiment 2], https://osf.io/prj8c/ [Experiment 3] https://osf.io/7uzet/ [Experiment 4], https://osf.io/kaqrn/ [Experiment 5], https://osf.io/f5hu7/ [Experiment 6], https://osf.io/3nzp8/ [Experiment 7], https://osf.io/bve4n/ [Experiment 8], https://osf.io/kqjfs/ [Experiment 9]). Importantly, because each study was based on a pilot study, we were able to preregister not just a general description of the statistical analyses used to test each hypothesis, but commented R syntax that had been tested on the pilot data. This preregistered syntax was then used to run the final statistical analysis, for which the data as well as the full analysis code is available in the by-experiment folders of the main OSF link [https://osf.io/7f3hd/].

2. Experiments 1-3: Reflexive Pronouns (*himself*): Principle A

2.1. Method

2.1.1. Participants

All participants were adult monolingual English speakers. The recruitment, inclusion and exclusion criteria are provided in full detail within the supplementary document [https://osf.io/7f3hd/]. In brief, for each of the main reflexive pronoun-interpretation studies (Experiments 1-3), we based our sample size on a power calculation from a pilot study (N = 10 in each case), stipulating a *minimum* power of 80% for each effect of interest (in practice, *a priori* power almost always cleared 90%). We additionally stipulated that the final N for each study must meet Brysbaert and Stevens' (2018) rule-of-thumb of 1600 observations per categorical condition level (full detail in the supplementary document). The final sample size for each of the main reflexive pronoun-interpretation studies was N = 134 for Experiment 1, N = 200 for Experiment 2, and N = 70 for Experiment 3. Each study also included a subsidiary event-likelihood rating task (see the supplementary document), with N = 50 (different participants to the main task).

2.1.2. Design

Experiments 1-3 investigated speakers' interpretation of sentences with reflexive pronouns (always *himself*[9]). On each trial, participants read a sentence of the form *Samuel asked Oliver about himself* (with different names counterbalanced across experimental conditions) and were asked to rate – using a continuous visual analogue scale – the extent to which *himself means*…: *Definitely Samuel*, *Definitely Oliver* or *Could equally be either* (see Figure 1).

[FIGURE 1 HERE]

As noted above (see also Endnote 1[A]), traditionally, interpretation of a sentence with a reflexive pronoun such as *Samuel asked/told Oliver about himself* has been determined by a syntactic constraint (Binding Principle A), such that the reflexive pronoun (e.g., *himself*) can refer only to *Samuel* (modified versions allow either the *Samuel* or *Oliver* interpretation, but do not attempt to explain which is preferred in each case[6]).

The aim of Experiments 1-3 was instead to investigate the extent to which participants' interpretations (ranging from SUBJECT [*Samuel*] to OBJECT [*Oliver*]) can be explained by a set of functionalist-pragmatic factors (summarized in Figure 2). Although, on the surface, these factors might seem diverse, all are simply different ways of operationalizing a single underlying construct: who the speaker most plausibly meant by *himself*. Sometimes, this is because the speaker has made a particular linguistic choice that directly reflects their communicative intentions. For example, if a speaker says *Someone* (c.f., Samuel) *told Oliver about himself*, then the speaker clearly intends the listener to infer that the identity of the teller is unknown and/or unimportant (i.e., low in the ***referential-hierarchy***; see Arnold et al., 2013; Gundel et al., 1993; Kuno, 1987; Van Valin & LaPolla, 1997); thus, *himself* probably refers to *Oliver*. But sometimes the construct of "who the speaker most plausibly meant by *himself"* is determined by reference to the real world (i.e, by ***event-likelihood***). For example, *Samuel* is more likely to ***tell*** *Oliver* about *Samuel*, than to ***tell*** *Oliver* about *Oliver*. Thus, if a speaker says *Samuel told Oliver about himself*, then – all other things being equal – they most plausibly meant *himself* to refer to *Samuel*, simply because that is the more likely real-world scenario (though as noted above, these underlying motivations have likely become crystalized into proceduralized interpretation shortcuts linked to individual verbs, nouns etc.). These functionalist-pragmatic factors for Experiments 1-3 are described in detail below.

[FIGURE 2 HERE]

2.1.2.1. Event-likelihood: Information source (verb manipulation, Experiments 1-3). People typically have access to much more information about themselves than about others. Thus, *Samuel* is more likely to **tell** *Oliver* about *Samuel*, than to **tell** *Oliver* about *Oliver* (i.e., *Samuel told Oliver about himself* [=*Samuel*]). Conversely, *Samuel* is more likely to **ask** *Oliver* about *Oliver* than to **ask** *Oliver* about *Samuel* (i.e., *Samuel asked Oliver about himself* [=*Oliver*]). We tested this prediction by (a) manipulating the verb (12 *tell*-type: *advise, alert, warn, inform, tell, notify, phone, mislead, correct, teach, email, text*; 12 *ask*-type: *ask, grill, interrogate, question, quiz, scrutinise, cross-question, challenge, interview, harass, hound, pester*); and (b) confirming the relative likelihood of SUBJECT versus OBJECT scenarios via a subsidiary rating task of sentences that did not include pronouns with separate participants (full details available in the supplementary document [https://osf.io/7f3hd/]). The event-likelihood manipulation is inspired by the study of Kaiser et al. (2009) who found a small but significant decrease in SUBJECT (*Samuel*) interpretations for "picture NP" sentences with *tell* versus *hear from* (e.g., *Samuel told/heard from Oliver about the picture of himself on the wall*).

2.1.2.2. Event-likelihood: Power relations (noun manipulation, Experiment 3). In a similar vein, given two specific characters and a verb, the odds on who will be the one asked about (or the one told about), are often far from 50/50. For example, a *lawyer* is more likely to interview a *suspect* about *the suspect* (e.g., "Where were you at 8pm on 15th October?") than about *the lawyer* (e.g., "How has my defence been so far?") (i.e., *The lawyer grilled the suspect about himself* [=*the suspect*]). Conversely, a *suspect* is more likely to tell a *lawyer* about *the suspect* (e.g., "I was at home, as my wife can testify") than about *the lawyer* ("You're a great lawyer!") (i.e., *The suspect told the lawyer about himself* [=*the suspect*]). In general, the more authoritative member of the pair is more likely to demand information about the other, who is more likely to provide it. We tested this prediction by (a) manipulating the noun pairs: *lawyer-suspect/suspect-lawyer, dad-boy/boy-dad, headteacher-pupil/pupil-headteacher, manager-employee/employee-manager*; and (b) confirming these aforementioned event-likelihood intuitions via a subsidiary rating task with separate participants (see the supplementary document: https://osf.io/7f3hd/).

2.1.2.3. Topicality: Referential-hierarchy (referring-expression manipulation, Experiment 1). Topicality refers to who (or what) a sentence is primarily about (Arnold et al., 2013; Kuno, 1987). Consider a sentence like ***Someone*** *shot the President*. The speaker has used an indefinite pronoun (*someone*) in the SUBJECT position, to indicate that – in the context of this conversation – the shooter is unknown and/or unimportant (i.e., is low on *topicality*). An indefinite noun phrase (***A man*** *shot the President*) and – more so – a definite noun phrase (***The man*** *shot the President*) each increases the emphasis on the shooter, but in each case the sentence is still very much about the President. A proper name (***Lee Harvey Oswald*** *shot the President)* places the shooter and the President on equal footing. However, a pronoun tips the balance (***He*** *shot the President*): Now the sentence is mainly about *he* [who we were already talking about]. The reversal (as compared with *Someone shot the President*) is complete: the sentence is firmly about the shooter (it could appear, for example, in a biography of the shooter). That is, the shooter is very high on *topicality*. The same is true if we manipulate the OBJECT position: *Lee Harvey Oswald shot* ***someone*** is about *Lee Harvey Oswald*; *Lee Harvey Oswald shot* ***him*** is about *the President* (or whoever we were just talking about). In general, then, given two phrases of this type (or, as they are called in the technical linguistics literature, "referring expressions"), the one that is higher in the following referential-hierarchy is the one that the sentence is primarily about and – crucially for our purposes – the one that *himself* is most likely to refer to:

Pronoun > Proper name > Definite Noun Phrase > Indefinite NP > Indefinite Pronoun
(e.g., *He/him > Samuel*/*Oliver > The man > A man > Someone*)

We tested this prediction by replacing either the SUBJECT (e.g., *Samuel*) or the OBJECT (e.g., *Oliver*), never both, with one of the five "referring expressions" from the hierarchy above (i.e., *Samuel asked [him / Oliver / the man /a man / someone] about himself; [He / Samuel / the man / a man / someone] asked Oliver about himself*).

2.1.2.4. Topicality: Prior-mention (context manipulation, Experiments 2-3). As we have just seen, pronouns are used when the character that the pronoun refers to is the topic of the conversation, and therefore – all other things being equal – the most likely referent of *himself* (e.g., *He told Oliver about himself* [*himself = He,* not *Oliver*]). The topicality of a pronoun is further boosted if the relevant character has just been mentioned in the context of a previous sentence (Harris & Bates, 2002; Hartshorne et al., 2015; Hendricks et al., 2013; van Rij et al.,

2011). For example, consider the same sentence with prior context for that SUBJECT expression *He:*

Samuel opened the door and stepped into the office.
He told Oliver about himself.

We are now close to certain that *himself* must mean *he* (i.e., refer to *Samuel*). Conversely, if we instead replace the OBJECT (*Oliver*) with a pronoun (*him*) and mention *Oliver* in a preceding context sentence, the interpretation of *himself* shifts to *Oliver*:

Oliver opened the door and stepped into the office.
Samuel told him about himself.

Although the use of *tell* (as opposed to *ask*) still pulls for a SUBJECT (*Samuel*) reading, the prior-mention of Oliver pulls strongly for an OBJECT (*him*; i.e., *Oliver*) reading, meaning we are much closer to 50/50. In general, then, pronominalizing a character (i.e., replacing his name with a pronoun) *and* mentioning him in a preceding context sentence increases the probability that *himself* refers to that character – independent of whether that character is in the SUBJECT or OBJECT position of the test sentence. We tested this prediction in both Experiments 1 and 2[10] by including prior-mention of either the SUBJECT (*Samuel*) or the OBJECT (*Oliver*) on every trial. The confirmatory prediction was that *himself* is more likely to be interpreted as the OBJECT when the OBJECT (rather than the SUBJECT) is topicalised by prior-mention in the context.

2.1.3. Stimuli and materials

Experiment 1 crossed the predictors of ***event-likelihood*** (24 continuous values corresponding to the 24 verbs), ***referential-hierarchy*** (5 categorical levels: proper name [e.g., *Samuel*], pronoun [e.g., *he*], other definite NP [e.g., *The man*], indefinite NP [e.g., *a man*], or indefinite pronoun [e.g., *someone*]), and ***position*** (binary levels: whether the grammatical SUBJECT or OBJECT is occupied by the *referential-hierarchy* manipulation). This yielded a total of 216 unique trials; though each individual participant completed only 108 (via counterbalanced lists which each randomly selected 12 out of the 24 verbs; always 6 *ask*-type and 6 *tell*-type).[11]

Experiment 2 crossed the predictors of ***event-likelihood*** (24 continuous values corresponding to the 24 verbs, as used for Experiment 1) and ***prior-mention*** (binary: whether the SUBJECT or OBJECT of the target sentence was topicalised by being mentioned in the context sentence). A further preregistered exploratory condition designed to de-confound prior-mention and pronominalization (*he/him*)[12] doubled the materials to yield a total of 96 unique trials, with each participant completing 48 (using the same counterbalanced-list procedure as Experiment 1).

Experiment 3 crossed the predictors of ***event-likelihood*** (48 continuous values as explained below) and ***prior-mention*** (binary: as above). This yielded a total of 96 unique trials, with each participant completing 48 (using the same counterbalanced-list procedure as Experiments 1-2). The ***event-likelihood*** predictor had 48 continuous values (rather than 24 in Experiments 1-2), because we crossed the verb (*N*=24) with the power-relation order (*N*=2; teacher-pupil vs pupil-teacher). For example, a difference score value was calculated from subsidiary study participants' ratings of the event-likelihood of *The headteacher asked the pupil about the headteacher* versus *The headteacher asked the pupil about the pupil*, another difference score was calculated from ratings of *The pupil asked the headteacher about the headteacher* versus *The pupil asked the headteacher about the pupil*; and likewise for all verbs and all power-relation pairs.

Procedure

The reflexive pronoun interpretation task was programmed and run using the web-based platform Gorilla.sc (Anwyl-Irvine et al., 2020). Participants completed the experiment remotely using their own computers or tablets in a single online session[13]. First, participants completed a 6-trial practice task designed to familiarize them with the procedure, but to give no clues as to the ratings expected in the main task [see the supplementary document: https://osf.io/7f3hd/]. For each trial of the main reflexive-pronoun (*himself*) interpretation task, participants saw a target sentence of the form *Samuel asked Oliver about himself* and were asked to rate – using a continuous visual analogue scale – the extent to which *himself means*…, *Definitely Samuel*, *Definitely Oliver* or *Could equally be either* (see Figure 1). For the statistical analyses, these analogue ratings were transformed into a 100-point scale (0 = *Definitely Samuel*, 50 = *Could equally be either*, 100 = *Definitely Oliver*), but no numerical values were ever shown to participants.

For Experiment 1, participants saw the target sentence only, presented in written form [publicly available to run here: app.gorilla.sc/openmaterials/875822]. For Experiments 2 [app.gorilla.sc/openmaterials/875823] and 3 [app.gorilla.sc/openmaterials/875824], participants additionally saw and heard a written and narrated prior-mention context sentence, which was followed by the text-only appearance of the target containing the pronoun. The addition of the audio-recorded sentences was designed to ensure that participants did not ignore the context sentence. After the target sentence appeared (e.g., *Samuel asked Oliver about himself*), participants responded to the prompt *himself means...* by using the visual-analogue scale, which formed the response variable for each Experiment.

For each study, separate participants also completed an event-likelihood rating task that asked about the likelihood not of sentence interpretations, but of real-world events (e.g., *How likely is it that Samuel asked Oliver about Samuel?* [SUBJECT focussed]; *How likely is it that Samuel asked Oliver about Oliver* [OBJECT focussed])?). We then subtracted each SUBJECT-focussed score from the corresponding OBJECT-focussed score to yield a difference score that operationalizes the relatively likelihood that, for example, *Samuel asked Oliver about Oliver*, rather than about *Samuel*. Full details are given in the supplementary document [https://osf.io/7f3hd/].

2.1.5. Analyses

A series of Generalised Linear Mixed-effects models (GLMMs) (Baayen, Davidson, & Bates, 2008) were fitted to the data in the R statistics environment (R Core Team, 2022) using lmer from the lme4 package (Bates et al., 2014). In each case, the dependent variable was the rating given to each pronoun-interpretation item on the 100-point scale (0 = *Definitely SUBJECT* [e.g., *Samuel*], 50 = *Could equally be either*, 100 = *Definitely OBJECT* [e.g., *Oliver*]). The independent variables were those set out in the *Stimuli and materials* section above, included both as main effects and – where relevant theoretical predictions could be derived – interactions. All models also included the continuous control predictors of verb frequency (log transformed) and trial order. In order to allow for meaningful interpretation of main effects in the presence of interactions, continuous predictors were centred, and binary predictors sum-coded. The discrete predictor of *referential-hierarchy* position (10 levels; see top panel of Figure 2) was Helmert-coded, yielding eight helmert-coded output terms (see the

supplementary document: https://osf.io/7f3hd/). The significance level was set at $p < .05$ in all preregistration documents, with *p* values calculated using the Kenward-Roger method (lmerTest package; Kuznetsova et al., 2017).

We followed the recommendations of Barr et al. (2013; also see Matuschek et al., 2017), by incorporating as many of these predictors as random slopes as were found to be warranted, removing each one if a Likelihood Ratio Test (LRT) indicated that its inclusion did not significantly improve model fit (i.e., $p > 0.05$), or if its inclusion led to convergence failure. Full details of the statistical approach, along with all data and analysis scripts can be found in the R scripts titled "MAIN"[14] in each subfolder of the main OSF link: https://osf.io/7f3hd/).

2.2. Results (Experiments 1-3)

In the interests of conciseness, we report here only the effects that relate directly to the preregistered predictions derived from the generalized functionalist-pragmatic account under investigation; all intended to operationalize the construct of who a speaker producing the relevant utterance would most plausibly intend by *himself*. Full models, with values for all fixed and random effects, can be found in the supplementary online material (see the R markdown script titled "entire_modelling" in each by-experiment subfolder of the main OSF link [https://osf.io/7f3hd/]).

2.2.1. Research Question 1. Does event-likelihood via information source (plus power-relations in Experiment 3) drive interpretation of reflexive pronouns (Experiments 1, 2 and 3)?

Our pre-registered confirmatory prediction was of a positive continuous relationship between the relative *event-likelihood* of OBJECT scenarios (positive difference score; e.g., *Samuel asked Oliver about* ***Oliver***) versus SUBJECT scenarios (negative difference score; e.g., *Samuel asked Oliver about* ***Samuel***) and participant interpretations of *himself* in sentences of the form *Samuel asked Oliver about himself* (0: *himself = Samuel*; 100: *himself = Oliver*). The confirmatory prediction was supported as a main effect in Experiment 1 ($b = 14.37$, $SE = 1.63$, $p < .001$, CI [11.17 – 17.57]; see Figure 3 x-axis), Experiment 2 ($b = 15.07$, $SE = 1.67$, $p < .001$, CI [11.79 – 18.34]; see Figure 4 x-axis), and Experiment 3 ($b = 6.24$, $SE = 0.81$, $p < .001$, CI [4.66 – 7.82]; see Figure 5 x-axis and by-shape for the distribution of Experiment 3's unique power-relation manipulation). Thus, in Experiment 1 for example, *himself* is

generally interpreted as meaning the OBJECT *Dave* in *Tom* ***asked*** *Dave about himself* (*M*= 80.69, *SE*= 3.55), but as meaning the SUBJECT *Tom* in the (structurally identical) sentence *Tom* ***misled*** *Dave about himself* (*M*= 17.49, *SE*= 2.51). Further confirmatory support was provided by Experiment 3 such that, further to the term reported above being significant at our predetermined alpha of $p <. 05$ (for which the predictor had 48 values corresponding to the 24 verbs presented in two power-relation orders), it also significantly improved the fit of the model relative to a model with only the 24-value likelihood predictor that corresponded to the 24 verbs ($\chi 2[2] = 202.12$, $p <.001$). That is, Experiment 3 similarly reported that *ask*-type verbs bias *himself* as being interpreted as the OBJECT; but, crucially, this was particularly in plausible scenarios where the OBJECT is the (non-authoritative) character who is more likely to provide information (e.g., *The suspect checked the time and then strolled into the room. The lawyer asked him about himself* [*M*= 79.92, *SE*= 4.89]), than when it is the (authoritative) character that would actually be more likely to demand information about the other (e.g., *The suspect checked the time and then strolled into the room. He asked the lawyer about himself* [*M*= 52.21, *SE*= 9.47]). Likewise a *tell*-type verb's bias for a SUBJECT interpretation was less pronounced when it was the less plausible (authoritative) character (most likely to demand information; e.g., *The suspect checked the time and then strolled into the room. The lawyer misled him about himself* [*M*= 52.03, *SE*= 9.78]) than when it was the plausible (non-authoritative) character who is likely to provide information (e.g., *The suspect checked the time and then strolled into the room. He misled the lawyer about himself* [*M*= 18.92, *SE*= 7.29]).

[FIGURE 3 HERE]

[FIGURE 4 HERE]

2.2.2. Research Question 2. Does Topicality via referential-hierarchy form drive interpretation of reflexive pronouns (Experiment 1)?

Figure 3 (by-colour referential conditions aligned to Figure 2) presents the mean interpretation responses for Experiment 1 in which referential-hierarchy was manipulated in the SUBJECT (top panel) or OBJECT position (bottom panel) (i.e., *Samuel asked [him / Oliver / the man / a man / someone] about himself; [He / Samuel / The man / A man / someone] asked Oliver about himself*). In line with our pre-registered confirmatory prediction, relative position in the referential-hierarchy of the SUBJECT and OBJECT significantly influenced responses. According to our preregistration, confirmatory support for

this prediction required a significant main effect for *only at least one* of the eight (Helmert-coded) referential-hierarchy output terms[15]. In fact, all eight terms were significant. For example, *himself* was generally interpreted as *John* in *John taught someone about himself* (*M*= 34.57**,** *SE*= 5.53) but as *Bill* in *Someone taught Bill about himself* (*M*= 59.75; *SE*= 4.72).

[FIGURE 5 HERE]

2.2.3. Research Question 3. Does topicality via prior-mention in context drive interpretation of reflexive pronouns (Experiments 2 and 3)?

Figure 4 (Experiment 2) and Figure 5 (Experiment 3) present the mean interpretation responses for each (by-colour) prior-mention condition. In line with our pre-registered confirmatory prediction, there was a significant effect of prior-mention, such that participants showed a greater preference for an OBJECT (blue) interpretation when it was topicalised by prior-mention, relative to when the SUBJECT (red) was topicalised by prior-mention (Experiment 2: b = -3.91, SE= 0.65, p < .001, CIs (-5.20 – -2.63); Experiment 3 [b = 2.63, SE= 0.98, p < .01, CIs [-4.55 – -0.71]). For example, *himself* was interpreted as the SUBJECT (*John, he*) in *John paused for a moment to take his jumper off. He warned Bill about himself* (*M*= 42.78, *SE*= 5.28), but (narrowly) as the OBJECT (*him, Bill*) in *Bill paused for a moment to take his jumper off. John warned him about himself* (*M*=52.35, *SE*=5.17). This effect also held in the exploratory analysis designed to de-confound pronominalization and prior-mention[16].

2.3. Discussion (Experiments 1-3)

Experiments 1-3 investigated the extent to which various broadly defined functional-pragmatic factors affect the extent to which participants interpret a reflexive pronoun (*himself*) as referring to the SUBJECT (e.g., *Samuel*) versus the OBJECT (e.g., *Oliver*) of sentences such as *Samuel emailed Oliver about himself*. All factors relate to the underlying construct of *who the speaker most plausibly meant by himself*. Experiments 1-3 yielded significant effects of ***event-likelihood*** (via information source) such that, for example, *himself* is generally interpreted as meaning *Oliver* in *Samuel asked Oliver about himself* but as meaning *Samuel* in the *Samuel misled Oliver about himself*. In a more fine-grained measure of *event-likelihood* (Experiment 3), OBJECT-biasing events like *asking* yielded even more OBJECT interpretations when the non-authoritative character was the OBJECT rather than

the SUBJECT and vice-versa. For example, *himself* is generally interpreted as *the pupil* in *The headteacher asked the pupil about himself*, but the OBJECT preference is alleviated in *The pupil asked the headteacher about himself.* Therefore, event-likelihood maps onto our single overarching construct of plausibility because it is a measure of how likely events are in the real world, for example: *How likely is it that person A would VERB person B about person A vs B?*

Experiment 1 yielded a significant effect of (topicality via) ***referential-hierarchy*** such that, for example, *himself* was generally interpreted as *Oliver* in *Someone emailed Oliver about himself* but as *Samuel* in *Samuel emailed someone about himself*. Experiments 2-3 yielded a significant effect of (topicality via) ***prior-mention*** in context such that, for example *himself* was generally interpreted as *he* (i.e., *Samuel*) in *Samuel opened the door and stepped into the office. He emailed Oliver about himself*; while *himself* was generally interpreted as *him* (i.e., *Oliver*) in *Oliver opened the door and stepped into the office. Samuel emailed him about himself.* These topicality manipulations function to make one character more accessible than the other, which therefore – all other things being equal – focus on this character as the most plausible referent of *himself,* as the topic of the conversation.

Crucially, as is clear from inspection of Figures 3-5, the size of these functional-pragmatic effects demonstrates that, across these sentences with *identical syntactic structure*, participants are not giving static structure-based interpretations (e.g., choosing 95% *Definitely Samuel* across all sentences) but instead graded interpretations that pattern according to the functional-pragmatic factors. Neither are these functional-pragmatic factors simply shifting participants' judgments a couple of percentage points in either direction; they are not merely the functionalist icing on the formalist cake. Rather, these functional-pragmatic factors can – while leaving the formal syntactic structure of the sentence identical – completely flip participants' interpretation of *himself* from SUBJECT (*Samuel*) to OBJECT (*Oliver*). To pick the most extreme examples (and focussing, for simplicity, on trials with no prior- mention in context), *himself* was generally (83%) interpreted as the SUBJECT (*Tom*) in *Tom misled someone about himself* (M = 17.25, SE= 3.24), but equally generally (85%) as the OBJECT (*him, i.e., Dave*) in *Tom asked him about himself* (M= 84.92, SE=2.36).

Having shown that these functional-pragmatic factors can together account for a considerable proportion of variance in participants' interpretations of reflexive pronouns (e.g., *himself*), we now ask whether they can do likewise for nonreflexive pronouns (e.g., *him*).

3. Experiments 4-6: Nonreflexive pronouns (*him*): Principle B

Experiments 4 to 6 investigated whether the functionalist-pragmatic account can also explain participants' interpretations of nonreflexive pronouns (e.g., *him*). Again, these experiments tested the functionalist-pragmatic prediction that, all else being equal, the pronoun refers to the most plausible referent, given the understood communicative intentions of the speaker. In broad terms, the design of the studies was analogous to those of Experiments 1-3, with the necessary adjustments described below.

The starting point for Experiments 4-6 was the well-studied phenomenon of so-called *picture Noun Phrases* (e.g., Jaeger et al., 2004; Keller & Asudeh, 2001; Runner et al, 2002). Simple sentences of the form used in Experiments 1-3 are not appropriate for investigating speakers' interpretations of nonreflexive pronouns (e.g., *Samuel asked Oliver about* ***him***), since functionalist (and, indeed, formalist) accounts make the prediction that *him* most naturally refers to *neither Samuel* nor *Oliver*, but to another character previously mentioned[17]. Picture noun phrases (e.g., *Samuel asked Oliver about the picture of him*) have received a good deal of attention in the literature because they constitute an apparent counterexample to the relevant formalist principle. Binding Principle B, at least in its original form (see Jaeger et al, 2004, for discussion of proposed modifications), stipulates that *him* cannot refer to either 'Samuel' or 'Oliver' (see Endnote 1[B]). Yet, as has long been noted (Jaeger et al., 2004), both interpretations (*him=Samuel*; *him=Oliver*) are possible for most English speakers. Focussing on these picture noun phrases, then, allows us to set aside formalist accounts, and – as for Experiments 1-3 – investigate the extent to which functional-pragmatic factors can explain the pattern of interpretations observed. That said, in order to check that our findings are not specific to picture noun phrases *per se*, we broadened the definition to include not just classic picture noun phrases (*picture of; photograph of*) but also *report*-type Noun phrases (*story about; news about*) and location noun phrases (*box next to; bag next to*).

3.1. Method

3.1.1. Participants

Criteria for participants and sample size were analogous to Experiments 1-3. The final sample size for each of studies 4-6 was $N = 70$ (with different participants in each study). As for Experiments 1-3, each study included a subsidiary event-likelihood rating task, with $N = 50$ (different) participants [see the supplementary document at https://osf.io/7f3hd/].

3.1.2. Design

In general, Experiments 4-6 followed the same design as Experiments 1-3, with two adjustments. First, since all sentences followed the template [*Samuel*] [*asked*] [*Oliver*] [*about*] *the* [*picture of*] *him*, the prompt was revised to *him means*… (rather than *himself*). Second, the response scale used the legends (for example) *Definitely Samuel* (left), *Could be Samuel* (centre) and *Definitely not Samuel* (right) (rather than, as for Experiments 1-3, *Could be either* and *Definitely Oliver*). This is because under traditional syntactic accounts, a nonreflexive pronoun (e.g., *him*) can in principle refer to anyone except the (local) SUBJECT (here, *Samuel*), not necessarily *Oliver* (the OBJECT). In practice, however, because only the SUBJECT and OBJECT are ever mentioned – these are the only realistic candidates for interpretation of *him*. For this reason, and for consistency with Experiments 1-3, we will still refer to choices as having an OBJECT interpretation. The functionalist-pragmatic factors were adapted as described below (see also Figure 6).

[FIGURE 6 HERE]

3.1.2.1. Event-likelihood: Information source and content (verb and prepositional phrase manipulation: Experiments 4-6).

As we saw for Experiments 1-3, events vary considerably as to their real-world likelihood. To take an extreme example, a lawyer is much more likely to question a suspect about a photograph of the suspect ("Doesn't the picture from the security camera show you stealing the money?") than about a photograph of the lawyer ("Wouldn't you agree that this suit brings out the colour of my eyes?"). Crucially, however, the real-world probability of these events varies not just with the verb type (*ask*-type/*tell*-type), but with the particular verb itself (*warn, tell, phone, email*; *ask, question, challenge, interview*) and the way it combines with the relevant prepositional phrase (*picture of, photograph of, story about, news about, box next to, bag next to*). For example, if *Samuel challenged Oliver about the news story about him*, the most natural interpretation is that *Samuel* is unhappy because *Oliver*, a journalist, has written an unfair news story about *Samuel* (i.e., *him=Samuel*). But if *Samuel warned Oliver about the picture of him*, the most natural interpretation is that a compromising photograph of *Oliver* is doing the rounds (i.e., *him=Oliver*).

We therefore continuously manipulated event-likelihood by crossing six prepositional phrases (*picture of, photograph of, story about, news about, box next to, bag next to*) and eight verbs (four *tell*-type: *warn, tell, phone, email*; four *ask*-type: *ask, question, challenge,*

interview). This formed 48 events, for which the relative likelihood of SUBJECT versus OBJECT (i.e., NON-SUBJECT) scenarios was calculated via a subsidiary rating task with separate participants (see the supplementary document [https://osf.io/7f3hd/]). Analogous to Experiments 1-3, the confirmatory prediction was simply that the more likely the event (in the subsidiary likelihood-rating task) the more likely the corresponding interpretation (in the main pronoun-interpretation task). Since each experiment crossed event-likelihood with at least one other categorical predictor (listed below), counterbalanced lists ensured that, although each participant saw each level of a categorical predictor appear with all eight verbs, that level and verb combination was only be seen with one version of the *photograph/picture of* synonym, one version of the *news/story about* synonym, and one version of the *box/bag next to* synonym (each determined by random allocation).

3.1.2.2. Event-likelihood: Power relations (noun manipulation, Experiment 6). Again, real-world plausibility varies with power relations. If *Samuel interviewed Oliver about the photograph of him*, *him* could plausibly refer to either *Samuel* or *Oliver* (or *someone* else altogether). But if *The lawyer interviewed the suspect about the photograph of him*, the power relations here are such that *him* almost certainly refers to *the suspect* ("Is that *you* pictured driving the getaway car?"). Using the same power-relations pairs as for Experiments 1-3, the confirmatory prediction was again simply that the more likely the event (in the subsidiary likelihood-rating task) the more likely the corresponding interpretation (in the main pronoun-interpretation task). Again, details of the supplementary rating task can be found in the supplementary document [https://osf.io/7f3hd/].

3.1.2.3. Topicality: Referential-hierarchy (referring-expression manipulation, Experiment 4). The five referential-hierarchy expressions, used in Experiment 1, were reduced to four to avoid ambiguous - and, without context, confusing – repetition of *him* (e.g., *Samuel asked him about the picture of him*). Otherwise, the prediction was the same as for Experiment 1, except referring to interpretation of *him*, rather than *himself*:

> Proper name > Definite Noun Phrase > Indefinite NP > Indefinite Pronoun
> (*Samuel*/*Oliver* > *The man* > *A man* > *Someone*)

Thus, for example, in both *Samuel emailed someone about the picture of him* and *Someone emailed Samuel about the picture of him*, *him* is (by hypothesis) more likely to refer to

S*amuel* (proper name) than s*omeone* (indefinite pronoun), even though *Samuel* is the SUBJECT in the first sentence and the OBJECT in the second.

3.1.2.4. Topicality: Prior-mention (context manipulation, Experiments 5-6). As for the reflexives in Experiments 2 and 3, the confirmatory prediction here was that *him* is more likely to be interpreted as the OBJECT when the OBJECT (rather than the SUBJECT) is topicalised by prior-mention in a suitable context (e.g., *Oliver opened the door and stepped into the office. Samuel asked Oliver about the picture of him*) and vice-versa.

3.1.3. Stimuli and materials

Experiment 4 crossed the predictors of ***event-likelihood*** (48 continuous values corresponding to the 8 verbs and 6 prepositional scenarios), ***referential-hierarchy*** (4 categorical levels: proper name [e.g., *Samuel*], other definite NP [e.g., *The man*], indefinite NP [e.g., *A man*], or indefinite pronoun [e.g., *Someone*]), and ***position*** (binary levels: whether the grammatical SUBJECT or OBJECT is occupied by the *referential-hierarchy* manipulation). This yielded a total of 336 unique trials; though each individual participant completed only 168 (via counterbalanced lists of synonyms as described earlier).

Experiment 5 crossed the predictors of ***event-likelihood*** (48 continuous values, as used for Experiment 4) and ***prior-mention*** *in context* (binary: whether the SUBJECT or OBJECT of the target sentence was topicalised by being mentioned in a preceding context)[12.] This yielded a total of 192 unique trials, with each participant completing 96 trials (using the same counterbalanced-list procedure as Experiment 4).

Experiment 6 crossed the predictors of ***event-likelihood*** (96 continuous values corresponding to the 48 values above crossed with the power-relations noun manipulation [*N*=2; e.g., *The lawyer interviewed the suspect about the photograph of him*; *The suspect interviewed the lawyer about the photograph of him*]), and *prior-mention* in context (binary: as above). This yielded a total of 192 unique trials, with each participant completing 96 (using the same counterbalanced-list procedure as Experiments 4-5).

3.2. Results (Experiments 4 to 6)

3.2.1. Research Question 1. Does event-likelihood (manipulated by information source and scenario; plus power-relations in Experiment 6) drive interpretation of nonreflexive pronouns (Experiments 4, 5 and 6)?

Analogous to Experiments 1-3, the data from Experiments 4-6 supported all our confirmatory predictions for the role of event-likelihood in interpretations of nonreflexive pronouns. The confirmatory prediction for event-likelihood was supported as a main effect in Experiment 4 [b = 3.55, SE= 0.66, p < .001, CI (2.25 – 4.85)]; see Figure 7 x-axis], Experiment 5 [b = 5.08, SE= 0.79 , p < .001, CI (3.52 – 6.64)]; see Figure 8 x-axis], and Experiment 6 [b = 7.35, SE= 0.63, p < .001, CI (6.12 – 8.58)]; see Figure 9 x-axis and by-shape for the distribution of Experiment 6's unique power-relation manipulation]. As for Experiments 1-3, event-likelihood was a very strong driver of interpretations. For example, in Experiment 4, *him* was generally interpreted as the OBJECT (*John*) for *Bill interviewed John about the picture of him* (M= 79.28, SE= 6.06), but as the SUBJECT (*Bill*) for *Bill phoned John about the picture of him* (M= 37.17, SE= 6.00).

The data from Experiment 6 additionally supported our confirmatory predictions regarding power relations. For example, *him* was more often interpreted as *the suspect* than *the lawyer*, regardless of whether it was in the SUBJECT position (e.g., *The suspect walked into the room and sat on a chair. He told the lawyer about the photograph of him* [M= 25.33, SE= 5.58]) or OBJECT position (e.g., *The suspect walked into the room and sat on a chair. The lawyer told him about the photograph of him* [M= 75.74, SE= 7.63]). Thus, incorporating power-relations into event-likelihood ratings (96 values) yielded a significantly better model fit than a model that excluded power-relations from event-likelihood ratings (48 values) [$\chi 2(2) = 609.44$, p <.001].

3.2.2. Research Question 2. Does topicality via referential-hierarchy form drive interpretation of nonreflexive pronouns (Experiment 4)?

Figure 7 (by-colour referential conditions, aligned to Figure 6) presents the mean interpretation responses for Experiment 4 in which position in the referential-hierarchy was manipulated in the SUBJECT (top panel) or OBJECT position (bottom panel) of the sentence e.g., (*Samuel asked [Oliver / the man / a man / someone] about the picture of him*); ([*Samuel / The man / A man / Someone] asked Oliver about the picture of him*]). Just as for interpretation of *himself* in Experiment 1, and in line with our pre-registered confirmatory prediction for Experiment 4, the referential-hierarchy position of the expression used for the SUBJECT and OBJECT significantly influenced interpretation of *him* in the predicted

direction, with all six Helmert-coded model terms significant. For example, for sentences with the verb *challenge*, *him* was overwhelming interpreted as SUBJECT (*Bill*) when he OBJECT was detopicalised [e.g., *Bill challenged someone about the photograph of him* (*M*= 29.67, *SE* = 6.98)], but as OBJECT (*John*) when the SUBJECT was detopicalised [e.g., *Someone challenged John about the photograph of him* (*M*= 68.21, *SE*= 6.01)].

3.2.3. Research Question 3. Does topicality via prior-mention in context sentence drive interpretation of nonreflexive pronouns (Experiments 5 and 6)?

Figure 8 (Experiment 5) and Figure 9 (Experiment 6) present the mean interpretation responses for each (by-colour) prior-mention condition. The data are again consistent with our pre-registered confirmatory prediction Experiment 5: b = -10.60, *SE*= 1.12, $p < .001$, CIs -12.79 – -8.42)[18]; Experiment 6 [b = -10.27, *SE*= 1.54, $p < .01$, CIs (-13.30 – -7.24)]. Specifically, participants generally interpreted *him* as OBJECT (e.g., *Dave*; shown in blue) when the OBJECT was topicalised by prior-mention (e.g., *Dave drank some coffee and then put it on the desk. Tom emailed him about the photograph of him* [*M*=85.83, *SE* = 5.56]), but as SUBJECT (e.g., *Tom*; red) when the SUBJECT was topicalised by prior-mention (e.g., *Tom drank some coffee and then put it on the desk. He emailed Dave about the photograph of him* [*M=38.33, SE = 12.30*]).

[FIGURE 7 HERE]

[FIGURE 8 HERE]

[FIGURE 9 HERE]

3.3. Discussion (Experiments 4-6)

As for Experiments 1-3, all preregistered predictions were confirmed, such that participants' interpretations of *him* (as for *himself* in Experiments 1-3) were affected by the relative real-world event-likelihood of the competing interpretations (as rated by independent participants), and topicality in terms of both referential-hierarchy (e.g., *Samuel* > *Someone*) and prior-mention. As for Experiments 1-3, it is not the case that these real-world semantic-pragmatic factors explained only a small ("icing on the cake") amount of variance. Rather, these factors, when combined, can – while holding syntax constant – yield either a strong (75%) SUBJECT (e.g., *Bill*) interpretation (e.g., *Bill emailed someone about the picture of*

him; *M*= 24.97, *SE*= 5.32) or a strong (85%) OBJECT (e.g., *Dave*) interpretation (e.g., *Someone told Dave about this picture of him*; *M* = 83.51, *SE* = 4.62).

4. Experiments 7-9: Pronouns in SUBJECT position (*He*): Principle C

For Experiments 7-9, we adapted the methodological approach used so far for reflexive and nonreflexive pronouns in an OBJECT position (e.g., *Samuel asked Oliver about* [***himself****/the picture of* ***him***], to investigate participants' interpretation of pronouns in a SUBJECT position (e.g., *he*). The functionalist-pragmatic explanation that we provided for these types of sentences in the context of example [c] *He kicked Samuel* (also see Ambridge et al., 2014; Kuno, 1987; Lakoff, 1968; MacWhinney, 2008, 2009) extends to an array of literature demonstrating that SUBJECT-position personal pronouns are interpreted as the current discourse topic: the "central character" in the unfolding narrative (e.g., Arnold, 2010; Grüter et al., 2018; Kehler et al., 2008; McDonald & MacWhinney, 1995; Pyykkönen & Järvikivi, 2010; Schumacher et al., 2017). An example is the well-studied phenomenon of implicit causality: in *Samuel criticised Oliver because he…*, *he* likely refers to *Oliver*, as the listener would expect to hear what *Oliver* did to prompt *Samuel's* criticism; whereas, in *Samuel amazed Oliver because he…*, *he* most plausibly refers to *Samuel*, as the listener would expect to learn what *Samuel* did to amaze *Oliver.*[19] However the overwhelming majority of prior studies (for the notable exception, see Harris & Bates, 2002) have examined cases like the above where *he* gets its meaning from a – typically previously mentioned[4] – foregrounded expression (an example applied to our specific stimuli would be our filler sentences like *Yusuf was driving home when he started indicating*[21]), for which functionalist accounts do not straightforwardly make predictions that go above and beyond a binding-principles account. That is, as far as the mainstream interpretation of Binding Principle C (see endnotes 1[C] and 7) would go for our implicit causality example above, *he* is not syntactically blocked from referring to *Samuel* nor *Oliver* (with that said, formalists [e.g., Reinhart, 1983:42] acknowledge the need for 'add-ons' to explain that the potential interpretations [i.e. *Samuel, Oliver*] do not occur in free variation, but vary systematically according to functionalist-pragmatic factors like the one we have offered above).

To address this issue, we specifically investigate whether functional-pragmatic factors can predict the circumstances under which *he* can refer to an entity mentioned in a later subordinate clause, using sentence forms like ***He*** *was driving home when Yusuf started indicating* (*he = Yusuf*). The single underlying theoretical construct remains unchanged, such that – by hypothesis – *he* is interpreted as referring to whoever the speaker most plausibly

means: either *Yusuf* or an alternative character. For the example above (***He*** *was driving home when Yusuf started indicating*) the alternative interpretation whereby *he* does not mean *Yusuf* may seem unlikely. However, for examples such as *He was driving home when Yusuf started cooking*, it is clear that the alternative interpretation, whereby *he* means '*someone other than Yusuf*', is much more likely (though again, these interpretations may be to some degree automized, rather than computed in real-time).

The starting point for Experiments 7-9 is the study of Harris and Bates (2002) in which participants – as in the present study – judged the interpretation of sentences such as *He was threatening to leave when Billy noticed that the computer had died*. For sentences of this type (the same form as used in the present Experiments 7-9), the initial main clause containing the pronoun *he* is backgrounded by aspectual markers of an ongoing "scene-setting" event, while the subordinate clause (*when…*) is foregrounded as the part of the sentence that drives the story forward. With this manipulation in place, participants judged *he* to refer to (e.g.,) *Billy* on a substantial majority of trials (85%) (Harris & Bates, 2002). With sentences of this type as our starting point (e.g., ***He*** *was driving home when Yusuf started indicating*, we again investigated the extent to which functional-pragmatic factors reflecting the speaker's inferred communicative intentions can explain participants' variance in interpretations of *he* from *Definitely Yusuf* to *Definitely NOT Yusuf*.

4.1. Method

4.1.1. Participants

Applying the same criteria as for Experiments 1-6, the final sample size for each of the main pronoun-interpretation studies was $N = 160$ for Experiment 7, $N = 160$ for Experiment 8, and $N = 54$ for Experiment 9. Experiment 9 also included a subsidiary rating task to operationalise the values of its continuous event-likelihood predictor (see the supplementary document [https://osf.io/7f3hd/], with $N = 50$ (different participants to the main task).

4.1.2. Design

Each of the trials followed a similar format to those of Experiments 1-6, but with two adjustments. First, all target sentences followed the form *He was [VERBing] when [Name] started [VERBing]* (e.g., *He was exercising in the gym when Yusuf started whistling*[20]). That is, the first clause always contained a pronoun in SUBJECT position and an ongoing, nonpunctual event (e.g., *He was driving home…*). The second clause always contained a full noun phrase in its SUBJECT position (*e.g., Yusuf / the man / a man / someone…*) and a

punctual event (e.g., *started indicating / burping / cooking*). Accordingly, the response scale was labelled *he means*…[9]. Second, the response scale used the legends (for example) *Definitely NOT Yusuf* (left), *Could be Yusuf* (centre) and *Definitely Yusuf* (right). This is because under traditional syntactic accounts, Binding Principle C BLOCKS a main-clause SUBJECT pronoun (*he*) from referring to a subsequent full noun phrase in a subordinate clause (…*when* ***Yusuf***…), but says nothing about who it CAN refer to. Accordingly, the present results are reported in terms of the extent to which participants allowed co-reference with the character explicitly mentioned in the target sentence (e.g., the extent to which *he* is interpreted as referring to *Yusuf*). The functionalist-pragmatic factors were as follows (also see Figure 10).

[FIGURE 10 HERE]

4.1.2.1. Event-likelihood: communication versus perception of punctual scenarios (binary verb class manipulation: Experiments 7-8). A sentence like *He was waiting in the office when Yusuf noticed that the paperwork had vanished* does not imply the presence of a second character in our mental representation of the scene. But if we replace *notice* (a perception verb) with *point out* (a communication verb), the presence of a second character is now implied (*He was waiting in the office, when Yusuf pointed out that the paperwork had vanished*), (if not, who could *Yusuf* have been talking to?). Therefore our confirmatory prediction was that participants will give more responses towards the *Definitely NOT Yusuf* end of the scale for communication verbs (*said, pointed out, revealed, complained, announced*) than perception verbs (*saw, noticed, spotted, realised, discovered*), since communication verbs – but not perception verbs – imply the presence of someone else other than *Yusuf*, that *he* could plausibly refer to. Note that this factor was tightly controlled so that only the verb following *Yusuf* (i.,e., in the subordinate *when* clause) was manipulated (5x communication; 5x perception): the events in both the first clause (e.g., waiting in the office) and the second clause (e.g., the paperwork had vanished) were designed to be neutral, and followed a counterbalancing scheme described in the stimuli section.

4.1.2.2. Event-likelihood: semantic coherence (continuous manipulation, Experiment 9). A unique aspect of event-likelihood manipulated in Experiment 9 (only) was the joint semantic coherence – and therefore real-world event-likelihood – of the events described in the first (main) and second (subordinate) clauses. A 30-value continuous predictor was formed by

creating 10 first-clause ongoing events (e.g., *He was driving home*) and following each, in the second clause, with an event that would be, broadly speaking, implausible / neutral / plausible for the first-clause character (*He*) to be doing simultaneously (…*when Yusuf started cooking / burping / indicating*). We then confirmed these intuitions via a subsidiary rating task with separate participants (see the supplementary document: https://osf.io/7f3hd/). Our confirmatory prediction was of a positive continuous relationship between plausibility in this subsidiary rating task (i.e., the extent to which a single person could plausibly be performing both events), and participant responses in the main pronoun-interpretation task (0 = definitely NOT the named character of the target sentence; 100 = definitely the named character of the target sentence).

4.1.2.3. Topicality: Referential-hierarchy (referring-expression manipulation, Experiment 7).

Analogous to Experiments 1 and 4, Experiment 7 manipulated topicality via referential-hierarchy. Specifically, the second clause of the target sentence contained one of four noun phrases in the familiar hierarchy (e.g., *He was waiting in the office, when [Yusuf > the man > a man >someone] [VERB]ed that the paperwork had vanished*). As for Experiments 1 and 4, the assumption was that the higher its position in the hierarchy (*Yusuf > the man > a man >someone*), the greater the extent that the speaker has deliberately foregrounded this noun-phrase as the intended referent of the pronoun (here, *he*; *himself/him* in Experiments 1 and 4). Thus, the confirmatory prediction was that the higher its position in the hierarchy, the greater the extent to which participants will rate *He* as referring to the character mentioned in the second clause.

4.1.2.4. Topicality: Prior-mention (context manipulation, Experiments 8-9). Similarly to Experiments 2-3 and 5-6, Experiments 8-9 added a prior-mention manipulation that either (a) explicitly named and topicalised (via first-mention) a plausible alternative referent for *He* (e.g., *Abdul visited the law firm with Yusuf*); or (b) included no such character (e.g., *It was Wednesday 2nd August*). Both contextual conditions were followed by the same filler sentence (e.g., *The law firm was cluttered and disorganised*), then one of the target sentences described above (*e.g., He was waiting in the office when Yusuf noticed that the paperwork had vanished*). The confirmatory prediction was that participants will rate *He* as referring to the character mentioned in the second clause (e.g., *Yusuf*) to a greater extent when no plausible alternative referent is given (e.g., *It was Wednesday 2nd August*), than when a

plausible alternative referent is named and topicalized by first mention (e.g., *Abdul visited the law firm with Yusuf*).

4.1.3. Stimuli and materials

Experiment 7 crossed the two predictors of ***event-likelihood*** [binary, corresponding to the verb classes of perception versus communication], and ***referential-hierarchy*** [4 categorical levels: proper name (e.g., *Yusuf*), other definite NP (e.g., *the man*), indefinite NP (e.g., *a man*), or indefinite pronoun (e.g., *someone*)]. For each of these eight conditions, each participant saw 10 items (determined by the selection of individual perception/communication verbs and neutral events), according to the counterbalancing scheme set out in the supplementary document [https://osf.io/7f3hd/]. This yielded a total of 80 experimental trials (e.g., *He was waiting in the office when Yusuf noticed that the paperwork had vanished*); though each individual participant completed 160 trials due to the addition of 80 filler items. These fillers (which were excluded from the statistical analysis) were designed to avoid a scenario by which *he=Yusuf* interpretations are (at least on our reading of Principle C[7], as discussed above), disallowed for all trials.[21]

Note that, for simplicity, the explanation above uses only stereotypically male names (e.g., *Yusuf*) with *He*. In fact, 50% of participants instead saw only (stereotypically) female names (e.g., *Amira*) with *She*. Each participant only ever saw a single proper name, which appeared 40 times with 10 verbs, 2 neutral event combinations, and 2 experimental versus filler items (for a total of 160 trials per participant).

Experiment 8 crossed the two binary predictors of ***event-likelihood*** (communication versus perception verbs) and ***prior-mention*** in context – or not – of a plausible alternative referent for *He* (e.g., *Abdul visited the law firm with Yusuf* vs. *It was Wednesday 2nd August*). As with Experiment 7, Experiment 8 included 10 items per categorical condition (2x verb class, 2x context). This yielded a total of 40 experimental trials, though each participant completed 120 trials due to the addition of 80 filler trials designed to rule out the potential confound of order-of-mention of the characters in the prior-mention context sentences.[22] Note that five neutral context settings (law firm, building, company HQ, library, laboratory) were designed for Experiment 8 to flexibly fit all the neutral two-clause events of the target sentence that were re-used from Experiment 7.

Experiment 9 crossed the predictors of ***event-likelihood****: semantic coherence* (30 continuous values; e.g., *He was driving home when Yusuf started indicating > burping >*

cooking) and ***prior-mention*** *in context* (e.g., *The date on **Abdul's** calendar showed Wednesday 3rd August* vs. *The date on **the** calendar showed Wednesday 3rd August*). Note that the context items were revised for Experiment 9 so that they could appear flexibly across the new events that were created for the 30 likelihood items (also see Figure 10). This yielded a total of 60 experimental trials (30 event-likelihood items each presented with two prior-mention context conditions), though each participant completed 180 trials due to the addition of 120 filler trials.[23]

4.2. Results (Experiments 7 to 9)

4.2.1. Research Question 1. Does event-likelihood via verb class (Experiments 7-8) and semantic coherence (Experiment 9) drive forward interpretation of subject (he/she) pronouns?

The confirmatory prediction for **event-likelihood** via verb-class manipulation was supported as a main effect in Experiment 7 (b = 6.76, SE= 0.89, p < .001, CI [5.02 – 8.50]; mean interpretation responses in Figure 11 x-axis) and Experiment 8 (b = 8.71, SE= 1.10, p < .001, CI [6.57 – 10.86]; mean interpretation responses in Figure 12 x-axis[24],[25]). For example, participants were considerably more likely to allow *he* to refer to *Yusuf* for *He was getting ready to leave when Yusuf **realised** that the book had disappeared* (M=75.94, SE = 3.57) than for *He was getting ready to leave when Yusuf **said** that the book had disappeared* (M=35.11, SE=8.34).

The confirmatory prediction for **event-likelihood** via semantic coherence was also supported as a main effect in Experiment 9 (b = 20.34, SE= 1.37, p < .001, CI [17.66 – 23.01]) with an effect that spanned almost the entire response scale (see Figure 13). For example, participants were considerably more likely to allow *he* to refer to *Isaac* for *The date on the calendar showed Tuesday 22nd March. He was driving home when Isaac started indicating* (M=75.32, SE=12.94) than for *The date on the calendar showed Tuesday 22nd March. He was driving home when Isaac started cooking* (M=1.32, SE=3.02).

FIGURE 11 HERE

FIGURE 12 HERE

4.2.2. Research Question 2. Does Topicality via referential-hierarchy form drive forward interpretation of subject (he/she) pronouns? (Experiment 7)?

For Experiment 7, all three Helmert-coded referential-hierarchy output terms were significant (see Figure 11: by-colour). For example, participants generally interpreted *he* as '*Yusuf*' in *He was working on the computer when **Yusuf** discovered that the paperwork had vanished* (*M*=83.44, *SE*=9.83), but generally interpreted *he* as 'NOT *Yusuf*' (i.e., *someone*) in *He was working on the computer when **someone** discovered that the paperwork had vanished* (*M*= 24.77, *SE*=11.02).

4.2.3. Research Question 3. Does topicality via prior-mention in a context sentence drive forward interpretation of subject (he/she) pronouns? (Experiments 8 and 9)?

Analogous to Experiments 2-3 and 5-6, our pre-registered confirmatory prediction for the role of prior-mention in context was confirmed for Experiment 8 (b = 22.84, SE= 1.84, p < .001, CIs [19.23 – 26.44]; see Figure 12: by-colour) and Experiment 9 (b = 3.98, SE= 1.66, p = .02, CIs [0.71 – 7.24]; see Figure 13: by-colour). For example, participants generally interpreted *he* as *Elijah* for *It was Thursday the 30th of September. The building was well furnished and comfortable. He was getting ready to leave when Elijah noticed that the heating had stopped working* (*M*=85.88, *SE*=5.23). In contrast, participants generally interpreted *he* as *Definitely NOT Elijah* for ***Noah** went to the building with Elijah. The building was well furnished and comfortable. He was getting ready to leave when Elijah noticed that the heating had stopped working* (*M*=33.63, *SE*=6.79).

[FIGURE 13 HERE]

4.3. Discussion (Experiments 7-9)

Analogous to Experiments 1-6, Experiments 7 to 9 yielded confirmatory findings for all our functional-pragmatic predictors: (a) event-likelihood via verb class (e.g., *He was waiting in the office when **Yusuf** [realised > pointed out] that the paperwork had vanished*), (b) event-likelihood via semantic coherence (e.g., *He was driving home when **Yusuf** started [indicating > burping > cooking]*) and topicality (e.g., [*It was Wednesday the 3rd of August > **Abdul** visited the law firm with Yusuf*]*… He was waiting in the office when Yusuf realised that the paperwork had vanished*. Again, it is certainly not the case the interpretation was driven primarily by syntactic factors, with these pragmatic factors merely tweaking these interpretations. Indeed, recall that – if our interpretation of Binding Principle C is correct[7] – it should *never* be possible for *he/she* to refer to the character mentioned in a later subordinate clause (…*when Yusuf*) at all. In fact, recall from the above examples that, in the most extreme cases, participants generally (around 85%) preferred the *Yusuf* interpretation for sentences

like *He was working on the computer when Yusuf discovered that the paperwork had vanished*, while overwhelming (98%) disallowing it for the syntactically-identical *The date on Alfie's calendar showed Tuesday 12th July. He was working in the laboratory when Yusuf started bowling* (*M*=2.05 *SE* = 2.85).

5. General Discussion

How do English-speakers interpret pronouns such as *himself* (Experiments 1-3), *him* (Experiments 4-6) and *he* (Experiments 7-9)? Since at least Chomsky (1981), the dominant answer has been that listeners determine possible and impossible interpretations using highly abstract rules that – at least under many versions of the theory – are "part of the innately specified Universal Grammar" (Crain et al., 2017, p. 127). In the present set of studies, we tested an alternative possibility: that listeners' interpretations are based instead on their functional-pragmatic understanding of *what the speaker most likely intended to convey*, given both the speaker's choice of words (e.g., *Someone* versus *a man* versus *Samuel*) and the listener's knowledge about the world (e.g., that a *lawyer* is more likely to grill a *suspect* about a picture of the *suspect* than about a picture of the *lawyer*). In order to isolate these functional-pragmatic factors, we devised – for each set of three studies – a series of sentences that hold syntactic structure constant (e.g., Experiments 1-3: *Samuel told Oliver about himself*; Experiments 4-6: *Samuel told Oliver about the picture of him*; Experiments 7-9: *He was driving home, when Yusuf started coughing*). This allowed us to set aside formalist accounts and investigate the extent to which functional-pragmatic factors alone can explain patterns of pronoun interpretation.

Across all nine studies, participants' judgments varied according to the relative real-world event-likelihood of the possible interpretations, to the speaker's choice of the particular words used to refer to the characters given considerations of topicality (referential-hierarchy), and to whether or not other characters had been previously mentioned. Crucially, these factors did not merely nudge participants' judgments a few percentage points in either direction. In all studies, these functional-pragmatic factors conspired to explain a range of judgments from around 85% SUBJECT to 85% OBJECT (or NOT SUBJECT), leaving very little variance unexplained.

To echo a point that we made in the Introduction, these findings do not – and, in principle, *cannot* – disprove the existence of possibly innate abstract syntactic binding

principles. But given the present evidence that a large proportion of the variability in participants' pronoun interpretation can be explained by functional-pragmatic factors, the onus is on those who would posit abstract binding principles to show exactly what additional findings or phenomena these principles explain. To clarify what we mean by "a large proportion", the mean pseudo conditional *r*-squared across all our models nears 50% (0.47), which is impressive given the various additional broader functional-pragmatic factors that come into play when we consider the wider cognitive framework that seeks to explain how a listener builds a mental model of the discourse (Johnson-Laird, 1983; Van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998). These include – to name but one of the factors left unexplored by the present research – framing and tracking entities from a certain viewpoint (MacWhinney, 2008, 2009).[26]

A caveat is in order here: We have, of course, shown that these functional-pragmatic factors explain substantial variability in participants' pronoun interpretations *only* for the three specific English sentence types studied. It remains possible that there are other sentence types for which these functional-pragmatic principles make exactly the wrong predictions, leaving syntactic binding principles to ride to the rescue. Nevertheless, a crucial lesson of the present series of studies is that any such claim will need to be investigated systematically and meticulously by carefully operationalizing and testing the relevant functional-pragmatic factors across hundreds of thousands of experimental trials, manipulating everything from the verb (e.g. *tell* vs. *ask*) and the verb's arguments (e.g., *someone+the man* vs. *the lawyer+the suspec*t), to the existence (or not) of a prior discourse context in which one of the characters is mentioned etc. It will not suffice simply to wave a couple of example sentences and assert – without conducing any kind of systematic investigation at all – that "functional-pragmatic factors can't explain *this*". Instead, we invite colleagues who believe that they have a phenomenon that constitutes a genuine counterexample to get in touch and arrange a collaboration (perhaps an "adversarial collaboration" in the sense of Clark & Tetlock, 2023), using methods along the lines of those set out in the present article. Maybe there really are cases of pronoun interpretation that are not explained by functional-pragmatic factors, but that can be well explained by (modified versions of) innate syntactic binding principles.

The present research joins a growing body of work which suggests that linguistic phenomena traditionally attributed to highly abstract (and possibly innate) grammatical principles can often fall naturally out of functional-pragmatic considerations. For example, Ambridge and Goldberg (2008) and Abeillé et al. (2020; and see Chaves & Putnam, 2021) provided evidence that the ungrammaticality of questions such as **Which sportscar did the*

color of delight the baseball player because of its surprising luminance? is the result of a pragmatic clash (essentially, "the colour of [say] the Porsche" is being treated as both backgrounded and foregrounded at the same time). Just as for the present study, these and similar restrictions had traditionally been explained as resulting from an innate formal syntactic principle; in this case, subjacency (e.g., Chomsky, 1986). Similarly, the principle of "structure dependence" has often been discussed as a "parade case" (Crain, 1991) of an innate syntactic constraint that has no functional motivation. In brief, the claim is that the principle blocks learners from hypothesising impossible rules such as "move the first auxiliary [here is] when forming a question from a statement".

*The boy who **is** smoking is crazy → **Is** the boy who smoking is crazy?*

Yet functionalist research (e.g., Ambridge, Rowland & Pine, 2008; Fitz & Chang, 2017; Ambridge, Rowland & Gummery, 2020) suggests that the absence of such errors again falls naturally out of functional considerations; in this case the fact that the complex noun phrase (e.g., *the boy who is smoking*) forms a coherent unit (e.g., it refers to an identifiable individual) and hence cannot be split up.

A potential advantage of these types of functionalist explanations over formalist ones (even when the two do an equally good job of explaining the data) is that they explain why these constraints exist in the first place. In an important sense, positing an innate syntactic principle is not an explanation, but an abdication of one. Why can't *Samuel kicked him* mean 'Samuel kicked himself'? The formalist answer is essentially "It just can't". A formal syntactic principle blocks this interpretation, but there's no apparent reason *why* we have this formal syntactic principle in the first place; it's just a quirk of evolution. The functionalist account instead offers answers to these types of questions framed in terms of the types of functional-pragmatic considerations that *language users will need to master anyway* in order to become effective speakers and listeners ("What did the speaker most likely mean by that?", "Why did they use the word *Samuel* instead of *someone*?"). We hope the present research will inspire colleagues to investigate empirically whether other phenomena often attributed to highly abstract linguistic constraints can be just as well or even better explained in terms of functional pragmatic factors.

In the meantime, while the linguistic pronoun wars are far from over, what the present findings have shown – as an absolute minimum – is that it is impossible to understand the totality of the linguistic facts regarding the possible and impossible interpretations of English

reflexive, nonreflexive and subject pronouns (e.g., *himself, him, he*) without taking into account functional-pragmatic considerations.

References

Abeillé, A., Hemforth, B., Winckel, E., & Gibson, E. (2020). Extraction from subjects: Differences in acceptability depend on the discourse function of the construction. *Cognition, 204*, 104293.

Adger, D., & Svenonius, P. (2015). Linguistic explanation and domain specialization: A case study in bound variable anaphora. *Frontiers in Psychology, 6*, 1421. https://doi.org/10.3389/fpsyg.2015.01421

Ambridge, B., Rowland, C. F., & Gummery, A. (2020). Teaching the unlearnable: A training study of complex yes/no questions. *Language and Cognition, 12*(2), 1–25. https://doi.org/10.1017/langcog.2020.5

Ambridge, B., & Lieven, E. V. M. (2011). *Child Language Acquisition: Contrasting theoretical approaches*. Cambridge, UK: Cambridge University Press.

Ambridge, B., & Goldberg, A. E. (2008). The island status of clausal complements: Evidence in favor of an information structure explanation. *Cognitive Linguistics, 19*(3), 349–381.

Ambridge, B., Pine, J. M., & Lieven, E. V. M. (2014). Child language acquisition: Why universal grammar doesn't help. *Language, 90*(3), e53–e90. http://dx.doi.org/10.1353/lan.2014.0051

Ambridge, B., Rowland, C. F., & Pine, J. M. (2008). Is structure dependence an innate constraint? New experimental evidence from children's complex question production. *Cognitive Science, 32*(1), 222–255.

Arnold, J. E. (2010). How speakers refer: the role of accessibility. *Language and Linguistics Compass, 4*, 187–203. 10.1111/j.1749-818X.2010.00193.x

Arnold, J. E., Kaiser, E., Kahn, J. M., & Kim, L. K. (2013). Information structure: Linguistic, cognitive, and processing approaches. *Wiley Interdisciplinary Reviews: Cognitive Science, 4*(4), 403–413. https://doi.org/10.1002/wcs.1234

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390–412.

Bates, D. M., Maechler, M., & Bolker, B. (2014). lme4: Linear mixed-effects models using S4 classes (R package version 0.999999–0). Retrieved July 28, 2023, from http://CRAN.R-project.org/package=lme4

Bergmann, C., Paulus, M., & Fikkert, P. (2012). Preschoolers' comprehension of pronouns and reflexives: The impact of the task. *Journal of Child Language, 39*(4), 777–803. https://doi.org/10.1017/S0305000911000298

Bickerton, D. (1975). Some assertions about presuppositions and pronominalizations. *Chicago Linguistic Society (Parasession on functionalism)*, *11*(2), 580–609.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition, 1*(1), 9. https://doi.org/10.5334/joc.10

Chaves, R. P., & Putnam, M. T. (2020). *Unbounded dependency constructions: Theoretical and experimental perspectives* (Vol. 10). Oxford Surveys in Syntax & Morphology, Oxford University Press.

Chien, Y.-C., & Wexler, K. (1990). Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition, 1*, 225–295.

Chomsky, N. (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, N. (1986). *Barriers*. Cambridge, MA: MIT Press.

Chow, W.-Y., Lewis, S., & Phillips, C. (2014). Immediate sensitivity to structural constraints in pronoun resolution. *Frontiers in Psychology, 5*, 630. https://doi.org/10.3389/fpsyg.2014.00630

Clackson, K., Felser, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language, 65*(2), 128–144. https://doi.org/10.1016/j.jml.2011.04.007

Clark, C. J., & Tetlock, P. E. (2023). Adversarial collaboration: The next science reform. In *Ideological and political bias in psychology: Nature, scope, and solutions* (pp. 905-927). Cham: Springer International Publishing.

Clifton, C., Frazier, L., & Deevy, P. (1999). Feature manipulation in sentence comprehension. *Italian Journal of Linguistics, 11*, 11–40.

Cole, P., Hermon, G, & Yanti (2015). Grammar of binding in the languages of the world: Innate or learned?. *Cognition*, *141*, 138-160.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences, 14*(4), 597–650.

Crain, S., Koring, L., & Thornton, R. (2017). Language acquisition from a biolinguistic perspective. *Neuroscience & Biobehavioral Reviews, 81*(Part B), 120–149. https://doi.org/10.1016/j.neubiorev.2016.09.004

Cunnings, I., Patterson, C., & Felser, C. (2014). Variable binding and coreference in sentence comprehension. *Journal of Memory and Language, 71*, 39–56. https://doi.org/10.1016/j.jml.2013.10.001

Dąbrowska, E. (2015). What exactly is Universal Grammar, and has anyone seen it? *Frontiers in Psychology, 6*, 852. https://doi.org/10.3389/fpsyg.2015.00852

Dillon, B., Mishler, A., Sloggett, S., & Phillips, C. (2013). Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language, 69*, 85–103. https://doi.org/10.1016/j.jml.2013.04.003

Fitz, H., & Chang, F. (2017). Meaningful questions: The acquisition of auxiliary inversion in a connectionist model of sentence production. *Cognition, 166*, 225–250.

Gordon, P. C., Grosz, B. J., & Gilliom, L. A. (1993). Pronouns, names, and the centering of attention in discourse. Cognitive Science, 17, 311–347. https://doi.org/10.1207/s15516709cog1703_1

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Speech acts*. New York, NY: Academic Press.

Grüter, T., Takeda, A., Rohde, H., and Schafer, A. J. (2018). Intersentential coreference expectations reflect mental models of events. *Cognition,* 177, 172–176. doi: 10.1016/j.cognition.2018.04.015

Gundel, J., Hedberg, N., & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language*, 69, 274–307.

Harris, C. L., & Bates, E. A. (2002). Clausal backgrounding and pronominal reference: A functionalist approach to c-command. *Language and Cognitive Processes, 17*(3), 237–269.

Harris, R. A. (1993). *The linguistics wars*. Oxford University Press.

Hartshorne, J.K., Nappa, R., & Snedeker, J. (2015). Development of the First-mention bias. *Journal of Child Language, 42*, 423-446. doi:10.1017/S0305000914000075

Hendriks, P., Koster, C., & Hoeks, J.C.J. (2014) Referential choice across the lifespan: why children and elderly adults produce ambiguous pronouns, *Language, Cognition and Neuroscience, 29:4*, 391-407.

Jackendoff, R. (1992). Mme. Tussaud meets the binding theory. *Natural Language & Linguistic Theory*, *10*(1), 1-31.

Jaeger, T. F., Butt, M., & King, T. H. (2004). Binding in picture NPs revisited: Evidence for a semantic principle of extended argument-hood. In *Proceedings of the LFG04 Conference*, Christchurch, New Zealand. Stanford: CSLI Publications.

Jäger, L. A., Mertzen, D., Van Dyke, J. A., & Vasishth, S. (2020). Interference patterns in subject-verb agreement and reflexives revisited: A large-sample study. *Journal of Memory and Language, 111*, 104063. https://doi.org/10.1016/j.jml.2019.104063

Järvikivi, J., van Gompel, R. P., & Hyönä, J. (2017). The interplay of implicit causality, structural heuristics, and anaphor type in ambiguous pronoun resolution. *Journal of psycholinguistic research, 46*, 525-550.

Kaiser, E., Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2009). Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition, 112*, 55–80.

Keller, F., & Asudeh, A. (2001). Constraints on linguistic coreference: Structural vs. pragmatic factors. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum.

Kehler, A., Kertz, L., Rohde, H., and Elman, J. L. (2008). Coherence and coreference revisited. *J. Semant.* 25, 1–44. doi: 10.1093/jos/ffm018

Kuijper, S. J. M., Hartman, C. A., & Hendriks, P. (2021). Children's pronoun interpretation problems are related to theory of mind and inhibition, but not working memory. *Frontiers in Psychology, 12*, Article 610401. https://doi.org/10.3389/fpsyg.2021.610401

Kuno, S. (1987). *Functional syntax: Anaphora, discourse, and empathy*. Chicago, IL: University of Chicago Press.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Levinson, S. C. (1987). Pragmatics and the grammar of anaphora. *Journal of Linguistics, 23*, 379–434.

Lidz, J., Lukyanenko, C., & Sutton, M. (2021). The hunt for structure-dependent interpretation: The case of Principle C. *Cognition, 213*, 104676. https://doi.org/10.1016/j.cognition.2021.104676

Macwhinney, B. (2005). The emergence of linguistic form in time. *Connection Science, 17*(3–4), 191–211. https://doi.org/10.1080/09540090500177687

MacWhinney, B. (2008). How mental models encode embodied linguistic perspectives. In R. L. Klatzky, B. MacWhinney, & M. Behrman (Eds.), *Embodiment, ego-space, and action* (pp. 369–409). Psychology Press.

MacWhinney, B. (2009). The emergence of grammar from perspective. In D. Pecher & R. A. Zwaan (Eds.), *Grounding cognition: The role of perception and action in memory, language, and thinking* (pp. 198–223). Cambridge University Press.

Matthews, D., Lieven, E., Theakston, A., & Tomasello, M. (2009). Pronoun co-referencing errors: Challenges for generativist and usage-based accounts. *Cognitive Linguistics, 20*(3), 599–626. https://doi.org/10.1515/COGL.2009.026

McKee, C., Nicol, J., & McDaniel, D. (1993). Children's application of binding during sentence processing. *Language and Cognitive Processes, 8*, 265–290.

McWhorter, J. (2024, June 27). *The Tiniest Words Generate the Biggest Uproar.* The New York Times. https://www.nytimes.com/2024/06/27/opinion/gender-neutral-pronouns-them.html

Newmeyer, F. J. (2010). Formalism and functionalism in linguistics. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(3), 301–307.

Nicol, J., & Swinney, D. (1989). The role of structure in coreference assignment during sentence comprehension. *Journal of Psycholinguistic Research, 18*, 5–19.

Parker, D., & Phillips, C. (2017). Reflexive attraction in comprehension is selective. *Journal of Memory and Language, 94*, 272–290. https://doi.org/10.1016/j.jml.2017.01.002

Pollard, C., & Sag, I. (1992). Anaphors in English and the Scope of Binding Theory. *Linguistic Inquiry*, *23*(2), 261–303.

Pyykkönen, P., & Järvikivi, J. (2010). Activation and persistence of implicit causality information in spoken language comprehension. *Experimental. Psychology, 57*, 5–16. doi: 10.1027/1618-3169/a000002

R Development Core Team. (2023). *R: A language and environment for statistical computing.* Austria: R Foundation for Statistical Computing Vienna. ISBN 3-900051-07-0.

Reinhart, T. (1983). *Anaphora and semantic interpretation.* London: Croom Helm.

Reinhart, T., & Reuland, E. (1993). Reflexivity. *Linguistic Inquiry, 24*(4), 657–720.

Reuland, E. (2011). *Anaphora and language design.* MIT Press.

Runner, J. T., Sussman, R. S., & Tanenhaus, M. K. (2002). Logophors in possessed picture noun phrases. In *WCCFL 21* (pp. 401–414).

Schnoebelen, T., & Kuperman, V. (2010). Using Amazon Mechanical Turk for linguistic research. *Psihologija, 43*(4), 441–464. https://doi.org/10.2298/PSI1004441S

Schumacher, P., Roberts, L., & Järvikivi, J. (2017). Agentivity drives real-time pronoun resolution: Evidence from German er and der. *Lingua* 185, 25–41. doi: 10.1016/j.lingua.2016.07.004

Thornton, R., & Wexler, K. (1999). *Principle B, VP ellipsis and interpretation in child grammar.* Cambridge, MA: MIT Press.

Truswell, R. (2014). Binding theory. In A. Carnie, Y. Sato, & D. Siddiqi (Eds.), *The Routledge Handbook of Syntax* (Vol. 1, pp. 214–238). (Routledge Handbooks in Linguistics). Taylor & Francis. https://doi.org/10.4324/9781315796604.ch11

Tucker, L., & Jones, J. (2023). Pronoun lists in profile bios display increased prevalence, systematic co-presence with other keywords and network tie clustering among US Twitter users 2015-2022. *Journal of Quantitative Description: Digital Media*, *3*.

van Hoek, K. (1995). *Anaphora and conceptual structure.* Chicago: University of Chicago Press.

van Rij, J., van Rijn, H., & Hendriks, P. (2013). How WM load influences linguistic processing in adults: A computational model of pronoun interpretation in discourse. *Topics in Cognitive Science, 5*(4), 564–580. https://doi.org/10.1111/tops.12029

Van Valin, R. D., Jr., & LaPolla, R. J. (1997). *Syntax: Structure, meaning, and function.* Cambridge: Cambridge University Press.

Xu, F. (2019). Towards a rational constructivist theory of cognitive development. *Psychological Review, 126*(6), 841–864. https://doi.org/10.1037/rev0000153

Yanti, Cole, P., & Hermon, G. (2017). The grammar of binding in the languages of the world: A response to Reuland. *Cognition, 168*, 380–384. https://doi.org/10.1016/j.cognition.2017.04.001

[1] In Lectures on government and binding, Chomsky (1981: 188) formulates these rules as follows [with our explanation of these rules as they apply to simple SUBJECT VERB OBJECT sentences like (a)-(c) in the main text]

(A) **An anaphor** [i.e., a reflexive pronoun; e.g., *himself*] **is bound in its governing category** [here, the governing category is the sentence; thus the reflexive pronoun is bound by – i.e., gets its meaning from – the sentence SUBJECT; here, *Samuel*].
(B) **A pronominal** [i.e., a nonreflexive pronoun; e.g. *him*] **is free in its governing category** [again, the governing category is the sentence; thus the nonreflexive pronoun is NOT bound by – i.e., does NOT get its meaning from – the sentence SUBJECT; here *Samuel*].
(C) **An R-expression** [i.e., a word that refers to something in the world, like *Samuel*, or *the man*] **is free** [meaning that it is NOT bound by – i.e., does not get its meaning from – the sentence SUBJECT]

[2] From the outset, we should clarify that we are not suggesting that listeners consciously reason through these processes for every use of every pronoun but, rather, that this reasoning has been routinised by repeated exposures across various contexts (e.g., Ambridge et al., 2014; Järvikivi et al., 2017; MacWhinney, 2008; McDonald & MacWhinney, 1995; Pyykkönen & Järvikivi, 2010). We thank an anonymous reviewer for raising this point.

[3] This additional factor of viewpoint – often referred to as the perspective hypothesis (MacWhinney, 2005, 2008) – is not specifically manipulated in the present studies, and is left as a further functional-pragmatic factor to be investigated in future research. Here, since *Samuel* is the starting point of the sentence, his viewpoint is initially adopted (see Gernsbacher, 1990). This viewpoint is maintained upon processing *himself* because it denotes the self-directed scenario described whereby *Samuel* maintains active control of the action and commands the viewpoint of our mental representation. That is, there is no other entity intervening to shift perspective away to an objective viewpoint; the event involves no other characters. Conversely, if the speaker intends to convey that *Samuel* is kicking another male character (as in [b]), the viewpoint of the representation shifts to an objective one that is no longer commanded by *Samuel*. Now, another male character is directly affected by the action, which necessitates the use of *him* (see MacWhinney 2005 and 2008 for examples of how refocusing devices can shift perspective away from the sentence-initial viewpoint, either leaving both *himself* and *him* possible, or framing them as mutually-exclusive possibilities). Also see endnote 26.

[4]Again, we thank an anonymous reviewer for raising this point. This principle (or the chain of reasoning that motivates it) also explains the interpretation of (Principle C) sentences of the form *Near John, he found a snake* (*he* $\neq$ *John*). Here, *he* cannot refer to *John*, even though *he* follows *John*. As discussed by Ambridge, Pine & Lieven (2014: 77):

> If a pronoun is used as the topic, this indicates that the referent is highly accessible, rendering anomalous the use of a full NP anywhere within the same clause (examples from Lakoff 1968, Kuno 1987).

a. *He$_i$ found a snake near John$_i$. (cf. John$_i$ found a snake near him$_i$.)
b. *Near John$_i$ he$_i$ found a snake. (cf. Near him$_i$ John$_i$ found a snake.)
c. *He$_i$ found a snake behind the girl John$_i$ was talking with. (cf. John$_i$ found a snake behind the girl he$_i$ was talking with.)
d. *He$_i$ loves John's$_i$ mother. (cf. John $_i$ loves his $_i$ mother.)
e. *John's$_i$ mother, he$_i$ adores dearly. (cf. His$_i$ mother, John$_i$ adores dearly.)

Of course, this can again be similarly formulated by a perspective hypothesis: the location information of the prepositional phrase acts to background *John* "so low in perspective that it cannot be a co-referent" (MacWhinney, 2009: 213). Conversely, the speaker could convey the intended meaning with *Near him, John found a snake* (*him* could mean *John*): now that the backgrounded location information is for the pronoun and the full name is foregrounded information, the preposition serves as a cue that signals that the listener can delay referential commitment.

[5] For example Cantrell (1974) offers "I can understand a father wanting his daughter to be like himself but I can't understand that ugly brute wanting his daughter to be like him". Under a formalist account the final *him* should really be *himself*, but – so the functionalist argument goes – *him* works better in this context as it gives the intended sense of the father being unattractive *from the speaker's point of view* rather than (necessarily) from the father's *own* point of view.

[6] As noted by an anonymous reviewer, some formalist accounts posit that a prepositional phrase constitutes an independent predicate to the verb, which allows a reflexive (e.g., *herself*) in an oblique argument to refer to the direct object argument (e.g., *herself* could mean 'Doris' or 'Mary' in *Mary explained Doris to herself*); see e.g., Sag and Pollard (1992); Jaeger et al. (2004). However, even with this modification, further functionalist-pragmatic 'add-ons' would be required to explain the fact that the two interpretations do not occur in free variation, but vary systematically according to the functionalist-pragmatic factors manipulated in Studies 1-3. Incidentally, Yanti et al. (2017) suggest that the "pure syntax" interpretation given here is indeed the most widespread in mainstream textbooks on syntactic theory.

7 An anonymous reviewer suggested the possibility that "the formalist account exempts application of Principle C outside of the clause". As far as we have been able to determine, this does not seem to be accurate (see e.g., Lidz et al, 2021, for a recent paper with examples of binding principles operating across clauses). Even if we are mistaken, however, exempting application of Principle C outside of the clause would serve only to make both interpretations (e.g., *Yusuf*; NOT *Yusuf*') possible, without making any predictions as to which will be preferred for a given sentence. The same can be said for the possibility (raised by a different anonymous reviewer) that the exclusive use of *when* in the test sentences might suggest a flat conjoined- ('and'), rather than subordinate-clause analysis, also exempting Principle C. Though, in any case, the conjoined-clause is unnatural for most of the test sentences, which do have a clear temporal sequence. For example, *He was waiting in the office when Yusuf noticed the paperwork had vanished* cannot easily be reformulated as *He was waiting in the office* ***and*** *Yusuf noticed the paperwork had vanished*, let alone *Yusuf noticed the paperwork had vanished* and *He was waiting in the office* (as would be possible for true conjoined clauses).

[8] We thank an anonymous reviewer for raising this point.

[9] Experiments 1-3 always used *himself* (and never *herself*), and stereotypically male names, for compatibility with Experiments 4-6, which always used *him* (and never *her*), on the basis that the pronoun *her* is at risk of confusion with possessive her (e.g., *her book*). However, Experiments 7-9 introduced sentences with a female pronoun (*she*) and stereotypically female names, finding an identical pattern of results to those observed with male names and pronouns (https://osf.io/7f3hd/ [see 'GenderCheck' markdown documents within 'EXPLORATORY' subfolder of the by-experiment folders]). Nongendered pronouns (e.g., *They asked them about themself*) were not used due to their relatively low frequency and ambiguous nature, which could have made it more difficult for participants to give reliable interpretations.

[10] An alternative would have been to compare context+target sentence pairs like *Samuel opened the door and stepped into the office + He asked Oliver about himself* (Experiments 2-3) with the equivalent no-context target sentences (*He asked Oliver about himself*) from Experiment 1. However, this would have left open the possibility that any difference observed could be due to the *mere presence* of a context sentence, regardless of its contents.

[11] Note that the 12 responses per participant for the proper name level were duplicated because it was the same sentence (e.g., *Samuel VERBed Oliver about himself*) regardless of whether the subject or object position was being replaced (i.e., N=108 responses per participant becomes N=120).

[12] Both Experiment 2 and Experiment 4 included an exploratory condition to further examine the proposed effects of topicality via prior-mention. Specifically, the character mentioned in the context sentence (e.g., *Samuel opened the door and stepped into the office*) was repeated in the target sentence as a proper name (i.e., *Samuel asked Oliver about himself*), rather than using a more natural pronoun expression (*He asked Oliver about himself,* as used for the main confirmatory prediction and our manuscript examples). The exploratory analysis is covered in the upcoming endnotes. Note that Experiments 3 and 6 investigated context effects using only the more natural pronoun expression, so did not include the exploratory condition.

[13] Various studies have demonstrated a high correlation between the online versus face-to-face delivery of linguistic-based experiments (e.g., Schnoebelen & Kuperman, 2010).

[14] Markdown files titled 'EXTRA' contain commentary for results that were also preregistered as being *confirmatory* to a hypothesis (often effects crossed within categories of other predictors), but are surplus to the main confirmatory effects that have been reported in the manuscript (typically collapsed over other predictors).

[15] This was simply out of a precautionary measure because we did not predict how the ratings might interact with potential floor and ceiling effects.

[16] In Experiment 2, an exploratory analysis revealed that significant context effects persisted even without the presence of the pronouns *he* or *him* (e.g., *Samuel/Oliver opened the door and stepped into the office. Samuel asked Oliver about himself*) [b = -1.92, SE= 0.86, p = .03, CIs (0.24 – 3.61)]. Whilst it is unnatural to use a proper name for a character's repeated mention (see Gordon et al., 1993), this exploratory finding of context effects for a context+target pair that uses full Noun Phrases (NOT pronouns) indicates that it occurs even

in the absence of a potential referential-hierarchy advantage that pronominalisation may have over proper names.

[17] The formalist account, for the reasons given in Endnote 1[B]; The functionalist account, on the basis of – amongst other things – the Gricean principle that "if the speaker meant 'himself', they would have said *himself*".

[18] Note that, unlike Experiment 2, this pattern did not hold in a preregistered exploratory check condition that manipulated prior-mention without also introducing pronouns [b = -0.81, SE = 1.12, p=0.47, CIs(-0.73–0.47)].

[19] This particular example draws from a phenomenon called implicit causality, which is compatible with our event-likelihood findings for OBJECT position pronouns in Experiments 4-6. It can be described as the expectations that listeners have for the coherence relations that hold between clauses (e.g. a listener's expectation for causal coherence, triggered by a verb that implies causality and/or the connective *because*, would in turn lead to an expectation that they will next hear about the cause of the event; see Kehler et al., 2008). In Experiments 4-6, our *asking* and *telling* verbs do not imply an expectation of causality, but rather imply that we will next hear about the source of knowledge (as we described in section 3.1.2.1). Crucially, it follows that well-established implicit causality effects on pronoun interpretation can theoretically fall under our single underlying construct of plausibility - after all, it is real world knowledge that drives implicit causality judgements (Pickering & Majid, 2007).

[20] A single proper name was used for all trials see by a given participant, and was randomly selected from a pool of 25: Tom, Joshua, Elijah, Oliver, Noah, Leo, Alfie, Arthur, Levi, Yusuf, Omar, Abdul, Isaac, Emma, Olivia, Katie, Chloe, Anna, Leah, Sophia, Maya, Amira, Jess, Aisha, Aaliyah. Note that, unlike in Experiments 1-6, we also introduced characters with stereotypically female names, and changed the subject pronoun from *he* to *she* accordingly.

[21] Specifically, these filler items for Experiments 7-9 flipped the clause order, such that the first clause was a subordinate clause (e.g., *When he was waiting in the office, Yusuf noticed that the paperwork had vanished*). Both Binding Principle C and functionalist-pragmatic factors (including simple order of mention) predict that *he= Yusuf* interpretations are possible here. An exploratory check of the descriptives was consistent with this: Experiment 7 (subordinate-main: M= 52.50, SE= 0.36; main-subordinate: M= 39.90, SE= 0.32); Experiment 8 (subordinate-main: M= 66.85, SE= 0.34; main-subordinate: M= 49.70, SE= 0.37); Experiment 9 (subordinate-main: M= 52.81, SE= 0.67; main-subordinate: M= 41.69, SE= 0.61).

[22] In Experiment 8, we first created 20 filler items akin to a third level to the context condition that was an exploratory check not part of the preregistered confirmatory hypothesis: this simply reversed the order of the names in the introduction sentence to topicalise the target sentence's second clause character (e.g., ***Yusuf*** *visited the law firm with Abdul*; see supplementary document). The 40 experimental items and 20 items for the third context level were then reused for all participants (to perform a total of 120 items) by flipping the clause-order, as above for Experiment 7. Note that the third context level served to disentangle two functional-pragmatic reasons (a no-explicit-competitor advantage versus a de-topicalisation effect) for the predicted confirmatory effect of context, which are explored and analysed in the supplementary material (https://osf.io/7f3hd/ see

'EXPLORATORY_PriorMention' markdown documents within 'EXPLORATORY' subfolder of the by-experiment folders).

[23] In the same vein as Experiment 8, Experiment 9 added a third level to the context condition as a pre-registered exploratory check. The third context level used the same character as in the second clause of the target sentence (no competitor character mentioned), e.g., *The date on **Yusuf's** calendar showed Wednesday 3rd August*. This yielded a total of 90 experimental trials doubled to 180 trials to swap the clause-order for the subordinate-main fillers (as above).

[24] Descriptive statistics of Experiment 7 confirm that the 5 verbs used for each class were performing as synonyms: communication verbs ranged from *M*= 35.07 to 37.48: *pointed out* (*M*= 35.07, *SE*= 0.67), *said* (*M*= 35.72, *SE*= 0.66), *announced* (*M*= 37.16, *SE*= 0.67), *complained* (*M*= 37.16, *SE*= 0.65), *revealed* (*M*= 37.48, *SE*= 0.67); perception verbs ranged from *M*= 41.83 to 44.92: *spotted* (*M*= 41.83, *SE*= 0.73), *discovered* (*M*= 42.27, *SE*= 0.73), *noticed* (*M*= 43.16, *SE*= 0.74), *saw* (*M*= 44.18, *SE*= 0.74), *realised* (*M*= 44.92, *SE*= 0.75).

[25] Descriptive statistics of Experiment 8 confirm that the 5 verbs used for each class performed consistently as synonyms: communication verbs ranged from *M* = 43.57 to 46.86: *said* (*M* = 43.57, *SE* = 1.01), *pointed out* (*M* = 43.82, *SE* = 1.02), *revealed* (*M* = 46.59, *SE* = 1.02), *announced* (*M* = 46.73, *SE* = 0.99), and *complained* (*M* = 46.86, *SE* = 0.99); perception verbs ranged from *M* = 52.71 to 56.26: *saw* (*M* = 52.71, *SE* = 1.00), *discovered* (*M* = 53.71, *SE* = 1.03), *spotted* (*M* = 54.02, *SE* = 1.01), *noticed* (*M* = 54.28, *SE* = 1.03), and *realised* (*M* = 56.26, *SE* = 1.03).

[26] Even though our manipulations can be conceptualised as altering the viewpoint/perspective (see Introduction and endnote 3) we did not manipulate perspective on a full scale per se. For example, in further ongoing experimental work we have shown that an obligatory use of *himself* when a listener represents *Samuel* entirely from his internal viewpoint (*Samuel splashed mud all over himself*) is substantially alleviated when a listener represents *Samuel* from an objective view from nowhere (*Samuel has mud all over him*).