

A probabilistic pragmatics for English singular *some**

Lewis Esposito
Stanford University

Christopher Potts
Stanford University

Abstract When English *some* combines with a singular NP, the resulting phrases reliably convey affective meanings not shared by variants with *a(n)* or plural NPs. Prior research has traced these effects to semantic properties of *some* that entail that the speaker cannot identify a unique referent for the phrase. In this paper, we present attested examples that conflict with this generalization. In addition, we argue that semantic accounts miss an important generalization: *some* is reliably affective only if *a* is available as an alternative. These facts suggest a pragmatic source for the relevant meanings. To capture them, we argue that a given context can make different *modes of identification* for entities relevant and that singular *some* signals a lack of engagement with these modes. We analyze the pragmatics of this signaling using the ‘lexical uncertainty’ version of the Rational Speech Acts model and show how it can be used to characterize the observed affective meanings.

Keywords: indefinites, English singular *some*, conceptual covers, Rational Speech Acts

1 Introduction

When English *some* combines with a singular NP, the resulting phrases often convey meanings not shared by variants with *a(n)* or plural NPs:

- (1) a. I met a friend.
b. I met some friends.
c. I met some friend.

Whereas (1a) and (1b) are affectively unmarked, (1c) is reliably (though not invariably) perceived as conveying *negative affective* meanings. Farkas (2002) and Weir (2012) trace these effects to semantic properties of *some* that entail that the speaker cannot identify a unique referent for the phrase. In essence, these accounts say that choosing *some* over its competitors is a choice to convey this uncertainty directly, which often leads to negative pragmatic inferences. Similar intuitions run

* Our thanks to Elisa Kreiss, Noah Goodman, Brandon Waldon, and participants at SALT 30 for their comments. Any errors are our own. The work is supported in part by a Google Faculty Research Award.

through semantic accounts of marked indefinites in other languages (Alonso-Ovalle & Menéndez-Benito 2003, 2010, 2011; Aloni & Port 2015; Martí 2015; Koev 2017; Solt & Stevens 2018).

In this paper, we identify two challenges for semantic accounts of English singular *some*. First, we present attested examples in which speakers use singular *some* to refer to entities that they themselves identify in the context (section 3). This argues against encoding speaker non-identifiability in the semantics. In addition, examples involving speaker-identifiable referents often remain affective, which suggests that non-identifiability cannot be the only source of the underlying meanings. Second, we argue that semantic accounts miss an important generalization: *some* conveys affective meaning only if *a* is available as competitor (section 3.3). Specifically, the meanings are absent when *some* combines with plurals and mass nouns, and they also seem to be bleached from words like *someone* and *something*, which show a different distribution from *a person* and *a thing*.

We argue that these challenges point to a pragmatic source for the relevant meanings. Our analysis builds on the general claim, explored in detail by Aloni (2000), that a given context can support many *modes of identification* for entities – for example, the ostensive mode, the naming mode, and descriptive modes that key into social roles and other aspects of identity. Our central claim is that a speaker who chooses singular *some* over *a* will convey that they are not engaging with any of the most relevant modes of identification. Aloni & Port (2015) reach a similar conclusion for Italian *un qualche* and German *irgendein*. This is the primary, stable pragmatic effect; more variable effects like negative affectivity are derivative of this – for example, some (but not all) contexts support an inference from lack of engagement with such modes to a lack of interest in them.

To make this account precise, we use (in section 4) the version of the Rational Speech Acts (RSA) model developed by Bergen, Levy & Goodman (2016), in which pragmatic reasoning leads synonymous forms that differ in markedness to take on different usage patterns. Our focus is on what messages signal about modes of identification. To bring these modes into the RSA model, we translate them into questions under discussion following procedures in Aloni 2000 and Aloni & Port 2015. The upshot is that singular *some* generally signals that the speaker is engaging only polar questions rather than more fine-grained questions tied to specific modes of identification. This association is the primitive pragmatic effect from which more variable and nuanced effects like negative affectivity emerge. Finally, in the RSA analysis, the effects depend on the presence of *a* as an alternative; if it is blocked for any reason, then *some* is predicted to lose its affectivity. This derives one of the core facts motivating a pragmatic account.

2 Some prior work

There is an extensive literature on the special meanings conveyed by indefinite forms in different languages. In the interest of space, we do not try to survey that literature here. Rather, we focus our discussion on the notion of *speaker non-identifiability* and various perspectives on it. These perspectives ultimately lead us to the idea that modes of identification, and context-dependent preferences around such modes, are important for understanding English singular *some*.

Farkas 2002 is a foundational theory of marked indefinite forms, covering a variety of languages, words, and morphosemantic contexts. For English singular *some NP*, Farkas proposes a discourse-semantic requirement that its associated discourse referent be *unidentified in context*: if a discourse referent x_i is introduced by *some NP*, then x_i must be compatible with multiple entities throughout the discourse. Examples like (2) are used to motivate this constraint:¹

- (2) Susan rented some movie for us yesterday.
 a. It was really long.
 b. #It was *The Maltese Falcon*. (Farkas 2002: 70)

In (2), *some movie* establishes a discourse referent x_i that could presumably pick out multiple actual movies. The continuation in (2a) is compatible with the non-identifiability constraint as long as it too leaves open the precise nature of the movie given the entities that the discourse participants take to be salient. In contrast, (2b) directly names the underlying referent, thereby clashing with the contextual restriction imposed by *some movie*.

Weir (2012) makes a related proposal. For Weir, singular *some* is similar to a regular indefinite semantically, but with an additional semantic presupposition that its first (domain) argument is true of more than one entity. This presupposition helps derive the *differentiation condition*, which says that a speaker uses *some* “to signal that she could not, if presented with the extension of NP, ‘pick out’ the witness of the existential claim”. This characterization is motivated by examples like (3):

- (3) You are lost. You know that the city you’re in has only two squares. You keep coming across both squares. You can tell them apart because one has a fountain and the other doesn’t. You end up in the fountainless square in the city. Your friend phones you:
 Friend: Where are you?
 You: I’m in a/#some square in the city. (Weir 2012: 182)

¹ Throughout this section, we cite examples from the literature with their published judgments. While the examples might in fact be marked, section 3 provides attested examples with the same general form, which challenges the notion that these judgments reflect general restrictions. Section 4.6 seeks to reconcile these intuitions.

For Weir, the *some* variant is marked because the speaker is in a context in which they can pick out the referent of *some square* in the relevant sense (even though their knowledge of that referent is partial and even falls short of what the context seems to require).

Solt & Stevens (2018) offer a related analysis of English *some*, focusing on combinations of *some* with numeral expressions. Building on Weir (2012), they posit an anti-singleton constraint on *some*, and they too predict that *some* is felicitous only if the NP is not identifiable for the speaker in context, again with important pragmatic qualifications as to what identification really entails.

Analyses of marked indefinites in other languages often center around notions of identifiability as well. The influential work of Alonso-Ovalle & Menéndez-Benito (2003, 2011) on the Spanish indefinite *algún* directly inspired the anti-singleton constraints of Weir (2012) and Solt & Stevens (2018). For these anti-singleton accounts, the idea is that inferences of speaker non-identifiability emerge as implicatures with respect to singleton-domain alternatives that the speaker could have uttered, but didn't. That is, listeners assume that a speaker using *algún* can't identify the referent of the NP, or else they would have chosen a different referring expression that doesn't have singleton domain restrictions.

However, examples like (4) (from Alonso-Ovalle & Menéndez-Benito 2011) show that we should be cautious when assuming that the usage patterns for these indefinites will be consistent across languages:

- (4) a. Look! Some professor's dancing on the table!
b. #¡Mira! Algún professor está bailando encima de la mesa.
Look! Some professor is dancing on the table!
Look! Some professor is dancing on the table.

While English *some* is compatible with ostension, Spanish *algún* is not. Examples like (4a) motivate Weir (2012) to develop more fine-grained ways of talking about identification and non-identifiability conditions.

A clear theme from the aforementioned work is that 'speaker identifiability' is itself a complex notion. Depending on the nature of the context, we might say that a referent is identifiable for a speaker in one context but not in another even if that speaker's knowledge is held constant. Aloni & Port (2015) place this kind of variability at the heart of their analysis of the Italian indefinite *un qualche* and the German indefinite *irgendein*. In essence, they argue that, when a speaker chooses *un qualche* or *irgendein*, they signal that they cannot identify the referent by means of the method of identification that is most relevant. For example, in a context in which the speaker needs to know the *name* of the referent in question, using one of these marked indefinites would signal that they lack this knowledge while allowing that they can identify the referent by other modes. In contrast, if the context

instead demands identification by a *professional role*, then these indefinites become consistent with identification by name (but not professional role).

In the next section, we offer a range of attested examples that help us distinguish among these ideas. To summarize briefly, the new examples seem to rule out very strict characterizations of non-identifiability; we see cases where singular *some* is used even where the speaker has command of all relevant notions of identifiability, a point of contrast with the Spanish, Italian, and German forms just discussed. Nonetheless, the examples strongly suggest that different modes of identification are important. This fact ultimately pushes us (in section 4) to characterize the usage conditions of *some* not in terms of what the speaker knows but rather in terms of what they aim to signal about the context.

3 Some data

This section seeks to characterize the core usage patterns for singular *some* that inform our analysis. We begin with a selection of attested examples, from the Corpus of Contemporary American English (CoCA; Davies 2008) and Web searches, that suggest a broader range of uses than previous literature has predicted. These examples begin to motivate our core characterization of singular *some* as signaling information about the speaker's intended modes of identification. Following that, we report on two more general distributional facts that help to motivate our analysis: a basic study of the frequency of singular *some* and a discussion of how *some*-based forms relate to (compete pragmatically with) other indefinites.

3.1 Some notable examples

Example (5) is a fairly typical instance of singular *some*, in which the speaker seems to be indicating that they lack relevant information about the referent:

- (5) Speaker A: Some lady called for you, she needs your help.
 Speaker B: What lady?
 Speaker A: I dunno.

There is certainly an intuition that it would seem inconsistent in some sense for Speaker A to reply to “What lady?” with “Your sister”, for example. This is an intuition that we would like to capture in our analysis. However, examples like the following begin to suggest that it would be too strong to derive this inconsistency from the semantics of *some*:

- (6) I saw some statue in Italy, and the expression on the guy's face blew me away. Like, I could read and recognize an emotion because some one 400

years ago felt how I do, and he made a stone the right shape so people could look at it and know exactly how he felt.

This example closely resembles (3). However, it seems clear that the speaker has access to a great deal of information about the referent of the singular *some* phrase. They could probably identify it by sight, and we don't sense any inconsistency if we assume that they could also identify the statue by name, by its location on a map, etc. Our own intuition about the example is that the speaker's choice to use *some statue* says more about what they take to be relevant than it does about their knowledge. Example (7) is similar:

(7) *Context: Rachel is checked into the motel and is standing next to the motel when she says:*

Rachel: To be honest, I came here to win someone back, and it didn't happen. So, my big grand gesture ends here, as I sleep alone in some roadside motel.

Here, the speaker *might* be able to name the motel, recite its address, etc., but *some* signals that these facts are not central, which helps imbue the story with a certain melancholy – the lost love makes the speaker's location beside the point.

That said, we might allow that the above examples involve only *partial* speaker identifiability. Examples like the following seem to show that *some* is consistent with full identifiability as well:

(8) Lorelei: It's embarrassing to ask my friend to turn around and get some kid at the middle school.

Patricia: What do you mean, some kid? He's your brother.

(9) Some guy I used to date texted me: "I miss spending time together".

In these examples, it is very clear that the speaker can identify the referent in many relevant ways – by name, by ostension, under numerous descriptions. We infer this from context and world knowledge in these cases, but such information can also be directly encoded in the phrases themselves, as we see in the following examples:

(10) The guitar player is some kid I know [...] and he shreds the guitar with reckless abandon and a creepy smile on his face. He looks like one of the members of Amebix, which fits perfectly with this band.

(11) Speaker A: Who's Tiffany Henderson?

Speaker B: Just some girl I used to date since I was like five. But that's over now. She blew me off.

Taken together, these examples indicate that speaker identifiability is not the central issue. What does seem to run through the examples is that they convey rich information about how relevant the speaker judges the identity of the referent to be. *Which*

statue is not relevant in (6) because just about any Italian statue will do. The identity of the hotel is beside the point in (7). In (8), Patricia seems taken aback by Lorelei's apparent lack of interest in the identity of the referent and so repeats "some kid" deliberately as a way of inquiring about this.

We also find examples in which the speaker uses both a proper name and singular *some* to refer to the same entity. In the following examples, the name is used first:

- (12) In Julian Assange's world we are the bad guys – not the Russians, not the Iranians, not the North Koreans. I hope the president-elect will get his information and trust the American patriots who work in the intelligence community who swear an oath and allegiance to the constitution and not some guy hiding from the law with a record of undercutting and undermining American democracy.
- (13) To borrow a phrase from Gloria Allred's favorite politician, Maureen Stenberg didn't build Staples. Tom Stenberg built Staples and is justly entitled to the earnings from Staples. If Romney somehow reduced the windfall that some woman got because her ex-husband founded Staples, I say good for him.

These examples are noteworthy in that the indefinite comes second but seems to corefer with the previously mentioned name. This is surprising given well-known preferences for indefinites to establish new discourse referents (Karttunen 1976; Kamp 1981; Heim 1982). However, if the two expressions involve non-trivial changes in the mode of identification, then it is easier to see why a speaker might opt for such a discourse structure (see also Heim 1998 on 'guises').

In (14) and (15), the indefinite is used before the name:

- (14) He's some creepy, rich overpaid football player, I mean come on [...] Michael Vick killed dogs, and he did [it] in a heartless and cruel way.
- (15) What does it mean for Kanye, who once sparked a very different sort of media moment by announcing on live television that "George Bush doesn't care about Black people" to appear in a photograph in a MAGA hat with some guy – namely music industry executive Lyor Cohen (more about him later) – flashing the OK gesture?

These examples might be the most challenging for accounts based on speaker non-identifiability, since both the proper names and the rich context suggest a great deal of speaker knowledge of these referents. If we instead adopt the view that singular *some* can convey information about modes of identification, then the examples make more intuitive sense: which modes are relevant can evolve as the discourse evolves, so that the proper name might be suitable at one stage and the indefinite at another.

In (16) and (18), we see that singular *some* can participate in discourse anaphora:

- (16) While I was at the airport, some relative of mine pointed at some people in the distance and whispered they were Syrians. I said ‘Syrians’ out loud a bit louder than I should have and the relative looked around to see if someone had overheard us. Yeah, the Syrians were kind of feared in Lebanon . . .
- (17) And some relative of mine has an amazing green house and a flower arranging place and she has a huge trash can there too [. . .] so I kinda based it off hers.
- (18) Fox News correspondent John Roberts, who was front and center at Wednesday’s press conference, joined *The Factor* with his takeaways from the bizarre event. “Jim Acosta of CNN was shut down a number of times,” Roberts said, “and it had become clear to me before the press conference that CNN would be iced out because they had published the report..”

Bill O’Reilly: But you have to know – you have to know – they could have said, listen, don’t worry about this. Because it comes from some guy we know who is smear merchant, has no credibility.

In (16) and (17), the singular *some* phrases establish new discourse referents that support subsequent definites. Example (18) is more involved: the discourse anaphoric link involves two speakers, and the indefinite appears to pick out a referent that is already established. As we noted in connection with (12) and (13), this is an unusual pattern for indefinites, but it would seem to make sense if the proper name and the indefinite involve different modes of identification.

3.2 Some corpus frequencies

The account we develop in section 4 depends on the claim that singular *some* is more marked than *a*. Our primary evidence for this comes from corpus frequencies. In a study of CoCA, we find that, while *some* across all of its uses is quite common, it is far less common than *a* when it modifies a singular NP. CoCA’s 2020 version contains roughly 1 billion words. There are 11,253,319 tokens of *a* followed by a singular noun phrase, and 433,109 tokens of *some* followed by a singular noun phrase. This means that *a* + singular NP is nearly 26 times more frequent than *some* + singular NP. This is a highly conservative estimate; singular *some* is likely even less frequent, as this number doesn’t exclude mass nouns (e.g., *some bread*).

3.3 Some alternatives

This brief section seeks to position singular *some* within the class of English indefinite expressions. The goal is to begin to map out the conditions under which a speaker can easily be said to be choosing *some* over alternatives. We identify two distributional patterns:

First, the indefinite article *a(n)* cannot take plural complements, so we can think of *some NP* as being pragmatically unopposed where *NP* is plural. It is tempting to cite bare plurals as alternatives, but Carlson (1977) argues extensively that this is incorrect based on contrasts in form and meaning. Similarly, other plural indefinites (e.g., *few NP*) carry different semantic denotations and don't always compete for the same syntactic slots as *some*. Plural *some* is also the more frequent context for *some*, occurring more than twice as frequently in CoCA as singular *some*. Second, *someone* and *something* are distributionally different from *a one/person* and *a thing*, so we claim that the relevant competition is not in play there either.

These facts relate to the core observation that plural *some* doesn't consistently give rise to the special affective meanings that we see with singular *some*, as we saw in (1), nor do forms like *someone* and *something*. If we were to propose a different semantic denotation for *some* with singular complements, we would need to carefully define which uses of *some* involved that meaning and ensure that no others could be interpreted as having that meaning. This would probably lead us to make distinctions that were motivated purely by the facts at hand. On the other hand, if we adopt a pragmatic account that depends crucially on speakers choosing *some* over alternatives, then we might be able to capture the distribution of these meanings in a uniform way without doing anything special in the semantics. We now turn to developing such a pragmatic account.

4 Probabilistic pragmatic account

The central generalization that emerges from the previous section is that singular *some* indicates that the speaker is not engaging with the most relevant modes of identification. The current section develops a formal account of this signaling.² The key ingredients are as follows. First, building on ideas from Aloni & Port (2015), we take all referential expressions to convey pragmatic information about the speaker's intended modes of identification, and we operationalize this notion using *conceptual covers* (Aloni 2000). Second, to make the connection between messages and modes of identification precise, we translate modes into *questions under discussion* (QUDs), again building on Aloni's (2000) analyses. Third, we apply the 'lexical uncertainty' version of RSA developed by Bergen, Goodman & Levy (2012) and Bergen et al. (2016),³ with special attention paid only to the cost function on messages to reflect

² Our OFS project <https://osf.io/3wqzc/> provides a complete implementation of this theory and includes all of the example simulations reported here.

³ The model of Kao, Bergen & Goodman (2014) is a potentially appealing alternative. That model defines a listener that reasons jointly about the world and the QUD. Kao et al. marginalize over possible QUDs, but one can as easily marginalize over worlds to capture the sense in which messages directly signal QUDs. We leave exploration of this alternative for future work.

W	[[the guitarist]]	[[the drummer]]	[[the singer]]	[[jesse]]	[[sandy]]	[[dani]]
w_1	j (signed)	s	d	j	s	d
w_2	j	s (signed)	d	j	s	d
w_3	j	s	d (signed)	j	s	d
w_4	j	s	d	j	s	d
w_5	d (signed)	j	s	j	s	d
w_6	d	j (signed)	s	j	s	d
w_7	d	j	s (signed)	j	s	d
w_8	d	j	s	j	s	d
w_9	s (signed)	d	j	j	s	d
w_{10}	s	d (signed)	j	j	s	d
w_{11}	s	d	j (signed)	j	s	d
w_{12}	s	d	j	j	s	d

Table 1 The intensional model for the simple scenario from section 4.1.

general and context-dependent aspects of markedness. The result of all this is that a speaker choosing singular *some* over *a* will be taken to be avoiding complex modes of identification in favor of simpler ones.

To try to make this analysis intuitive, we ground the discussion in a simple scenario (section 4.1) and an associated semantic theory (section 4.2). This is the foundation for our core message/QUD connection (section 4.3) and in turn for our pragmatic theory (section 4.4).

4.1 Simple scenario

We imagine that the speaker went to a concert, got their album signed by a band member, and returned home to say, “Some band member signed my album”. What does this convey beyond what we would get from alternatives like “A band member signed my album” or “X signed my album” for some band member’s name X?

To make the scenario concrete, imagine that the band has three members – Jesse, Sandy, and Dani. Our intensional model can be depicted as in table 1. The domain of entities is $D = \{j, d, s\}$ and the set of worlds is $W = \{w_1, \dots, w_{12}\}$. The columns in table 1 correspond to individual concepts and the ‘(signed)’ annotations indicate that that band member signed the speaker’s album in that world.

4.2 Semantics

This section builds up the semantic theory that we use for our pragmatic simulations. The key features: we rely on conceptual covers to define *who*-type questions (Aloni 2000), and we specify that *some* and *a* are semantically identical.

4.2.1 Individual concepts and conceptual covers

Names are rigid designators:

$$(19) \quad \llbracket \text{jesse} \rrbracket = \lambda w j \quad \llbracket \text{dani} \rrbracket = \lambda w d \quad \llbracket \text{sandy} \rrbracket = \lambda w s$$

In contrast, definite descriptions are non-rigid and follow the scenario. For example, this is the meaning of *the guitarist*, corresponding to the column under $\llbracket \text{the guitarist} \rrbracket$ in table 1:

$$(20) \quad \llbracket \text{the guitarist} \rrbracket = \lambda w \begin{cases} j & \text{if } w \in \{w_1, w_2, w_3, w_4\} \\ d & \text{if } w \in \{w_5, w_6, w_7, w_8\} \\ s & \text{if } w \in \{w_9, w_{10}, w_{11}, w_{12}\} \end{cases}$$

Following Aloni (2000), we capture the notion of a ‘mode of identification’ in terms of *conceptual covers*, which are sets of individual concepts. A set of individual concepts C is a conceptual cover for a domain D and set of worlds W iff

$$(21) \quad \forall x \in D \text{ and } \forall w \in W, !\exists c \in C : c(w) = x$$

Our scenario has exactly two covers:

$$(22) \quad \begin{aligned} \text{a. } \llbracket \text{name cover} \rrbracket &= \{ \llbracket \text{jesse} \rrbracket, \llbracket \text{sandy} \rrbracket, \llbracket \text{dani} \rrbracket \} \\ \text{b. } \llbracket \text{role cover} \rrbracket &= \{ \llbracket \text{the guitarist} \rrbracket, \llbracket \text{the singer} \rrbracket, \llbracket \text{the drummer} \rrbracket \} \end{aligned}$$

Compare these with $\{ \llbracket \text{the guitarist} \rrbracket, \llbracket \text{the drummer} \rrbracket \}$, which fails to be a cover because, for example, d has no associated individual concept in w_1 . Similarly, $\{ \llbracket \text{the singer} \rrbracket, \llbracket \text{the drummer} \rrbracket, \llbracket \text{jesse} \rrbracket \}$ isn’t a cover because j has two associated individual concepts in w_5 .

4.2.2 Predicates

Predicates are functions from entities into sets of worlds. The semantics of **signed** is as depicted in table 1. We also make use of a predicate **musician** that maps every entity in the domain D to the full set of worlds W – that is, our domain consists entirely of musicians.

For combining predicates with individual concepts, we can use this convenience function (‘intensional function application’):

$$(23) \quad \llbracket \text{ifa} \rrbracket = \lambda f \lambda c \{ w \in W : w \in f(c(w)) \}$$

This allows us to have expressions like $\llbracket \text{ifa}(\text{signed})(\text{jesse}) \rrbracket = \{w_1, w_6, w_{11}\}$ even though $\llbracket \text{signed} \rrbracket$ takes individuals as arguments and $\llbracket \text{jesse} \rrbracket$ is a function from worlds to entities.

English singular *some*

4.2.3 Quantifiers

Quantificational determiners are functions from pairs of predicates into sets of worlds. Here of course we are making the foundational claim that *some* and *a* are semantically identical:

$$(24) \quad \llbracket \mathbf{a} \rrbracket = \llbracket \mathbf{some} \rrbracket = \lambda f \lambda g \{w \in W : \exists x (w \in f(x) \cap g(x))\}$$

4.2.4 Interrogatives

We translate modes of identification to questions, which are derived from conceptual covers. We adopt the general tenets of the partition semantics for questions (Groenendijk & Stokhof 1984). In (25), we define polar interrogatives directly as mapping sets into sets of sets:

$$(25) \quad \llbracket ? \rrbracket = \lambda p \{p, W - p\}$$

With the above, we get questions like the following:

$$(26) \quad \llbracket ?\mathbf{jesse}(\mathbf{signed}) \rrbracket =$$

$$\left\{ \begin{array}{c} \{w_1, w_6, w_{11}\} \\ \{w_2, w_3, w_4, w_5, w_7, w_8, w_9, w_{10}, w_{12}\} \end{array} \right\} = \left\{ \begin{array}{c} \llbracket \mathbf{signed}(\mathbf{jesse}) \rrbracket \\ \llbracket \neg \mathbf{signed}(\mathbf{jesse}) \rrbracket \end{array} \right\}$$

$$(27) \quad \llbracket ?\mathbf{a}(\mathbf{musician})(\mathbf{signed}) \rrbracket =$$

$$\left\{ \begin{array}{c} \{w_1, w_2, w_3, w_5, w_6, w_7, w_9, w_{10}, w_{11}\} \\ \{w_4, w_8, w_{12}\} \end{array} \right\} = \left\{ \begin{array}{c} \llbracket \mathbf{a}(\mathbf{musician})(\mathbf{signed}) \rrbracket \\ \llbracket \neg \mathbf{a}(\mathbf{musician})(\mathbf{signed}) \rrbracket \end{array} \right\}$$

Our analysis of *who* questions is cover-sensitive in the sense of Aloni 2000. In the interest of readability, we define this as a function that outputs pairs of worlds that agree on the meaning of the predication defined by the conceptual cover and the complement clause:

$$(28) \quad \llbracket \mathbf{who} \rrbracket = \lambda C \lambda f \left\{ \begin{array}{c} \langle w, w' \rangle \in W \times W : \\ \left\{ \begin{array}{c} \{c \in C : w \in \llbracket \mathbf{ifa} \rrbracket(f)(c)\} \\ = \\ \{c \in C : w' \in \llbracket \mathbf{ifa} \rrbracket(f)(c)\} \end{array} \right\} \end{array} \right\}$$

It is then straightforward to map these sets of pairs of worlds into sets of sets, which will be equivalence classes of worlds defined by f and C .

These meanings are cover-sensitive in that we get different questions depending on which of the two covers we provide as the first argument. For example, if we use **name cover**, then we get a question that corresponds to the set of propositions determined by the name individual concepts (plus ‘no musician signed’).

$$(29) \quad \llbracket \mathbf{who}(\mathbf{name\ cover})(\mathbf{signed}) \rrbracket = \left\{ \begin{array}{l} \{w_1, w_6, w_{11}\} \\ \{w_2, w_7, w_9\} \\ \{w_3, w_5, w_{10}\} \\ \{w_4, w_8, w_{12}\} \end{array} \right\} = \left\{ \begin{array}{l} \llbracket \mathbf{ifa}(\mathbf{signed})(\mathbf{jesse}) \rrbracket \\ \llbracket \mathbf{ifa}(\mathbf{signed})(\mathbf{sandy}) \rrbracket \\ \llbracket \mathbf{ifa}(\mathbf{signed})(\mathbf{dani}) \rrbracket \\ \llbracket \neg \mathbf{a}(\mathbf{musician})(\mathbf{signed}) \rrbracket \end{array} \right\}$$

In contrast, if we use **role cover**, then we get the set of propositions corresponding to the description given by the roles:

$$(30) \quad \llbracket \mathbf{who}(\mathbf{role\ cover})(\mathbf{signed}) \rrbracket = \left\{ \begin{array}{l} \{w_1, w_5, w_9\} \\ \{w_2, w_6, w_{10}\} \\ \{w_3, w_7, w_{11}\} \\ \{w_4, w_8, w_{12}\} \end{array} \right\} = \left\{ \begin{array}{l} \llbracket \mathbf{ifa}(\mathbf{signed})(\mathbf{the\ guitarist}) \rrbracket \\ \llbracket \mathbf{ifa}(\mathbf{signed})(\mathbf{the\ singer}) \rrbracket \\ \llbracket \mathbf{ifa}(\mathbf{signed})(\mathbf{the\ drummer}) \rrbracket \\ \llbracket \neg \mathbf{a}(\mathbf{musician})(\mathbf{signed}) \rrbracket \end{array} \right\}$$

In what follows, we rely on the association that this creates between conceptual covers and questions. For example, we can talk about engaging with the name-based mode of identification or, equivalently, about the question in (29).

4.3 Questions as states

Normally in RSA analyses, the starting point is a lexicon that associates messages with worlds/states according to the semantics. However, we primarily want to characterize what messages signal about modes of identification. In casting these as conceptual covers, we can use our denotation of *who* to get from covers to questions as modes of identification. Once we have done this, the goal is to systematically relate messages to questions. Our proposed method for doing this employs a simple and familiar notion of answerhood:

$$(31) \quad \text{A proposition } p \subseteq W \text{ answers a question } q \subseteq \wp(W) \text{ iff there is a set } x \subseteq q \text{ such that } \bigcup_x = p.$$

Here $\wp(W)$ is the powerset of W . The definition just says that p answers q iff p is equivalent to a union of some cells in q . This is essentially the notion of *partial answer* from Groenendijk & Stokhof 1984 (and see Groenendijk 1999 for a closely related set of pragmatic definitions).

Putting our semantics together with this ‘Question as states’ view, we derive table 2 as the basis for RSA analyses in our scenario. Here, 1 means ‘answers’ and 0 means ‘doesn’t answer’, according to (31). To save space, we use $\llbracket ?\mathbf{jesse} \rrbracket$ to abbreviate (26), $\llbracket ?\mathbf{a} \rrbracket$ to abbreviate (27), $\llbracket \mathbf{who\ name} \rrbracket$ to abbreviate (29), and so forth for the others. Thus, the above specifies a set of states containing (i) all the

	[[who name]]	[[who role]]	[[?a]]	[[?jesse]]	[[?sandy]]	[[?dani]]	[[?guitarist]]	[[?singer]]	[[?drummer]]
some	1	1	1	0	0	0	0	0	0
a	1	1	1	0	0	0	0	0	0
jesse	1	0	0	1	0	0	0	0	0
sandy	1	0	0	0	1	0	0	0	0
dani	1	0	0	0	0	1	0	0	0
the guitarist	0	1	0	0	0	0	1	0	0
the singer	0	1	0	0	0	0	0	1	0
the drummer	0	1	0	0	0	0	0	0	1

Table 2 The answerhood relation A from (31).

polar questions we can create based on individual concepts and quantifiers as well as (ii) the two more fine-grained questions we can derive using the two conceptual covers in our scenario. As we would expect given that *some* and *a* are semantically identical, they also answer exactly the same questions.

4.4 Pragmatic model

Our pragmatic model is identical to the one developed by Bergen et al. (2016), which has been used to analyze a wide range of pragmatic phenomena (Potts & Levy 2015; Potts, Lassiter, Levy & Frank 2016; Champollion, Alsop & Grosu 2019; for a related proposal, see Jäger 2012). Our only modification is that we apply the model to message/question connections rather than message/state connections.

We can think of the model as having two layers. In the first layer, the nature of the message/question connections is fixed to be a single such structure A , as in table 2. Speaker and listener agents reason in terms of this fixed structure. In the second phase, the speaker and listener agents reason about *sets* of these structures all at once, which can be taken to represent their uncertainty about how exactly messages relate to (engage with) QUDs. This more abstract form of reasoning is key to capturing the effects of markedness where synonymous signals are involved.

With this in mind, let’s develop the first layer of the model. A *context* for this model is a tuple $(M, Q, A, P, Cost)$. Here M is a set of messages, Q is a set of possible questions under discussion, A is a ‘lexicon’ as in table 2, capturing the answerhood relation, P is a prior probability distribution over Q , and $Cost$ is a cost function on messages, with larger costs indicating more marked or costly messages.

The model begins with a literal listener l_0 :

(32)

$$l_0(q | m, A) = \frac{A(m, q) \cdot P(q)}{\sum_{q' \in Q} A(m, q') \cdot P(q')}$$

Given any message $m \in M$, this agent defined a probability distribution over ques-

tions, assigning positive probability only to questions q where $A(m, q) = 1$, and taking the prior probability of questions into account.

The first pragmatic speaker reasons about this literal listener:

(33)

$$s_1(m | q, A) = \frac{\exp(\alpha(\log l_0(q | m, A) - \text{Cost}(m)))}{\sum_{m' \in M} \exp(\alpha(\log l_0(q | m', A) - \text{Cost}(m')))}$$

This looks very involved but unpacks into just a few key pieces. On observing a state q , this agent defines a distribution over messages. It does this by considering the values assigned by l_0 while also taking the costs of those messages into account. We take the log of the l_0 values so that they are in a mathematical space in which it makes sense to subtract real-valued costs. The parameter α controls the strength of the inferences – larger α will increase the distances between the values calculated, thereby amplifying any pragmatic contrasts that they contain. Finally, the entire expression is exponentiated to take us from the log-valued space back into a space of values that can be normalized into a probability distribution.

In standard versions of RSA (e.g., Frank & Goodman 2012; Goodman & Stuhlmüller 2013), a pragmatic listener responds to this agent. This agent would have the same form as (32), but it would reason in terms of s_1 rather than the answerhood relation A . This model can generate a wide range of patterns that we might place under the heading of *scalar implicature*, but Bergen et al. (2016) show that it is not able to capture standard effects related to the markedness of messages.

To address this, Bergen et al. introduce the idea that the lexicon A is not fixed, but rather has some pragmatic uncertainty about it. For them, this uncertainty concerns the lexicon – or, more generally, the semantic system. In our application of these ideas, the uncertainty relates to the answerhood relationship.

To define this new model, we simply extend the above contexts into tuples $(M, Q, P, \text{Cost}, \mathbf{A}, P_{\mathbf{A}})$. The changes are that we don't have a fixed A , but rather a set of them \mathbf{A} , as well as a prior over those objects $P_{\mathbf{A}}$. The recursion suggested by (32) and (33) then continues with the following ‘uncertainty listener’:

(34)

$$L_1(q | m) = \frac{P(q) \sum_{A \in \mathbf{A}} s_1(m | q, A) P_{\mathbf{A}}(A)}{\sum_{q' \in \mathbf{Q}} P(q') \sum_{A \in \mathbf{A}} s_1(m | q', A) P_{\mathbf{A}}(A)}$$

This agent reasons about all of the A in \mathbf{A} . The precise nature of this set \mathbf{A} is left open by the model. Following Bergen et al., we take table 2 and allow every message to be refined so that it is compatible with any subset of the states it is compatible with in table 2. For example, since **jesse** is compatible with **[[who name]]** and **[[?jesse]]**, we allow lexicons where it is compatible with both of those as well as ones in which it is compatible with just **[[who name]]** and others where it is compatible with just

Cost		[[who name]]	[[who role]]	[[?a]]	[[?jesse]]	[[?sandy]]	[[?dani]]	[[?guitarist]]	[[?singer]]	[[?drummer]]
0.5	some	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0
0	a	0.5	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	jesse	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0
1	sandy	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0
1	dani	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
1	the guitarist	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0
1	the singer	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
1	the drummer	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0
	<i>P</i>	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11

Table 3 Primary simulation: L_4 with a available as an alternative. The costs of messages are given in the rightmost column, and the (flat) prior over questions is given in the bottom row. Gray highlights the most probable questions for each message.

Cost		[[who name]]	[[who role]]	[[?a]]	[[?jesse]]	[[?sandy]]	[[?dani]]	[[?guitarist]]	[[?singer]]	[[?drummer]]
0.5	some	0.33	0.33	0.33	0.0	0.0	0.0	0.0	0.0	0.0
1	jesse	0.00	0.00	0.00	1.0	0.0	0.0	0.0	0.0	0.0
1	sandy	0.00	0.00	0.00	0.0	1.0	0.0	0.0	0.0	0.0
1	dani	0.00	0.00	0.00	0.0	0.0	1.0	0.0	0.0	0.0
1	the guitarist	0.00	0.00	0.00	0.0	0.0	0.0	1.0	0.0	0.0
1	the singer	0.00	0.00	0.00	0.0	0.0	0.0	0.0	1.0	0.0
1	the drummer	0.00	0.00	0.00	0.0	0.0	0.0	0.0	0.0	1.0
	<i>P</i>	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11	0.11

Table 4 As in table 3 but with no a alternative.

[[?jesse]]. The set \mathbf{A} contains every combination of every lexical refinement. Given these assumptions, we obtain 35,721 lexicons from table 2. We always assume $P_{\mathbf{A}}$ assigns the same value to all of them.

From here, the model proceeds essentially as in the base case, but without the dependence on a particular A :

$$(35) \quad S_n(m | q) \propto \exp(\alpha (\log L_{n-1}(q | m) - \text{Cost}(m))) \quad n \geq 2$$

$$(36) \quad L_n(q | m) \propto S_n(m | q)P(q) \quad n \geq 2$$

To reduce clutter, we have left off the denominators here, specifying instead values that are proportional to the final values. To obtain the normalized values, one uses essentially the same denominator as in (33) for (35) and (32) for (36).

4.5 Simulations

Table 3 presents our primary simulation using the model just presented. For this simulation, we set $\alpha = 4$ and depict L_4 . These are aggressive pragmatic settings that

are informed by results in [Bergen et al. 2016](#) for capturing markedness implicatures. In addition, we need to make two specific assumptions about the cost function *Cost*.

First, we specify that *some* is more costly than *a*, reflecting the core markedness intuitions and the frequency imbalances discussed in section 3.2.

Second, to achieve the full range of desired effects, we must also specify that the other, more specific referring expressions (*Jesse, the guitarist, etc.*) are more costly than *some*. This is not necessarily motivated directly by our observations for *some*. Rather, we claim it is marked in general to try to engage very specific modes of identification using indefinites. For example, if the speaker wishes to engage with the mode of identification by name, then using a name is clearly better. For the scenarios we want to model, something must be preventing the speaker from doing this. The reasons could relate directly to which modes of identification are relevant. On the other hand, the reasons could be orthogonal to these considerations – for example, a concern that one of the discourse participants doesn't fully master these more specific forms. To set these issues aside, we assign these forms higher cost, with the result that *a* emerges as the best way of engaging these more specific questions, as desired. If we don't do this, then *some* and *a* both associate with $[[?a]]$.

In table 4, we simply remove *a* as an alternative and run exactly the simulation from table 3. This too shows the desired result: *some* is now equally likely to engage with the fine-grained 'who'-type questions and the polar question $[[?a]]$. This is the bleaching of the special meanings that we sought to predict in situations in which *some* is not perceived as a deliberate choice over *a*.

4.6 Discussion

The simulations achieve our central goal: we predict that *some* is a disfavored way of engaging with fine-grained modes of identification where *a* is available as an alternative. We now address important questions that this simulation raises.

First, we think our account can explain why the examples from section 2 might sound marked. We predict that the stable pragmatic effect of singular *some* is to convey that the speaker is not engaging with fine-grained modes of identification. In examples like (2b), the speaker transmits this signal and then, with no apparent motivation, directly engages with a precise mode of identification (naming). This switch is perhaps jarring enough that it comes across as truly infelicitous; see also the discussion in [Lauer 2013](#) of how uncancellable implicatures can arise in situations where cancellation cannot be motivated. We can compare this with examples like (12) and (13). These too use singular *some* and then subsequently name the entity. However, the speaker motivation is much clearer for these cases, and it seems we can even characterize it in terms of a desire to change the mode of identification.

Second, our account stops short of predicting negative affectivity directly. Our

claim is that negative affectivity is a derivative pragmatic effect rather than a primary one. What our analysis delivers is that a speaker using singular *some* is not engaging with fine-grained modes of identification. There can be many underlying reasons for this. Negative affectivity is likely to be strong where (i) contextual cues or world knowledge give us confidence that the speaker *can* engage with fine-grained modes of identification and (ii) those modes are relevant. In these contexts, if the speaker chooses *some*, they are deliberately avoiding relevant, pragmatically accessible information about a referent. This is where we get the surprised response from Lorelei in (8), and it contributes to the overall negativity of cases like (12).

Third, as we noted above, the decision to make *Jesse* and *the guitarist* and the others in that group more costly than *some* might seem unmotivated. Without this, both *a* and *some* act like one might expect a regular indefinite to act: they become associated pragmatically with $[[?a]]$. The names become preferred ways of engaging with $[[\mathbf{who\ name}]]$ and the definites are best for $[[\mathbf{who\ role}]]$. This seems natural given the pragmatic dynamic established by our system. That said, it does feel like a weakness of our account. One might hope that, conditional on hearing *a*, the listener would be open to associating it with the more specific modes. The deep pragmatic recursion we use to achieve our central results rules this out; if we instead look at L_1 rather than L_4 , holding everything else about our main simulation the same, we do see the desired pattern emerge, albeit weakly (e.g., $L_1([[mathbf{who\ name}]] | \text{some}) = 0.16$).

Finally, it is worth asking what happens if one specific mode of identification is more likely. Aloni & Port (2015) discuss situations in which fine-grained modes of identification differ in their relevance. We can model these differences using the prior term P . In particular, if we hold everything about the simulation in table 3 the same but allow $P([[mathbf{who\ name}]])$ to get very large, then the $L_4([[mathbf{who\ name}]] | a)$ cell goes towards 1, which is the manner implicature pattern studied by Bergen et al. (2016). This can be compared minimally with the corresponding values in table 3, where the probability mass for *a* is equally distributed across $[[\mathbf{who\ name}]]$ and $[[\mathbf{who\ role}]]$.

5 Conclusion

We presented a range of new attested examples of English singular *some* and argued that the common pragmatic effect running through them is that the speaker is signaling a lack of engagement with relevant modes of identification. Depending on the context, this might give rise to inferences that the speaker is ignorant about such modes, or doesn't regard them as relevant, which can result in negative affectivity. We used the 'lexical uncertainty' RSA model of Bergen et al. (2016) to capture these effects, which in effect assimilates them to manner implicatures, though with a focus on modes of identification rather than standard propositional content.

References

- Aloni, Maria. 2000. *Quantification under conceptual covers*. Amsterdam: Universiteit van Amsterdam PhD dissertation.
- Aloni, Maria & Angelika Port. 2015. Epistemic indefinites and methods of identification. In Luis Alonso-Ovalle & Paula Menéndez-Benito (eds.), *Epistemic indefinites: Exploring modality beyond the verbal domain*, 117–140. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780199665297.001.0001.
- Alonso-Ovalle, Luis & Paula Menéndez-Benito. 2003. Some epistemic indefinites. *North East Linguistic Society (NELS)* 33. 1–12.
- Alonso-Ovalle, Luis & Paula Menéndez-Benito. 2010. Modal indefinites. *Natural Language Semantics* 18(1). 1–31. doi:10.1007/s11050-009-9048-4.
- Alonso-Ovalle, Luis & Paula Menéndez-Benito. 2011. Domain restrictions, modal implicatures and plurality: Spanish *algunos*. *Journal of Semantics* 28(2). 211–240. doi:10.1007/s11050-009-9048-4.
- Bergen, Leon, Noah D. Goodman & Roger Levy. 2012. That’s what she (could have) said: How alternative utterances affect language use. *Cognitive Science Society (CSS)* 34. 120–125.
- Bergen, Leon, Roger Levy & Noah D. Goodman. 2016. Pragmatic reasoning through semantic inference. *Semantics and Pragmatics* 9(20). doi:10.3765/sp.9.20.
- Carlson, Gregory. 1977. *Reference to kinds in English*. Amherst, MA: UMass Amherst PhD dissertation.
- Champollion, Lucas, Anna Alsop & Ioana Grosu. 2019. Free choice disjunction as a rational speech act. *Semantics and Linguistic Theory (SALT)* 29. 238–257. doi:10.3765/salt.v29i0.4608.
- Davies, Mark. 2008. The Corpus of Contemporary American English. <http://corpus.byu.edu/coca/>.
- Farkas, Donka. 2002. Varieties of indefinites. *Semantics and Linguistic Theory (SALT)* 12. 59–83. doi:10.3765/salt.v12i0.2873.
- Frank, Michael C. & Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science* 336(6084). 998. doi:10.1126/science.1218633.
- Goodman, Noah D. & Andreas Stuhlmüller. 2013. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science* 5(1). 173–184. doi:10.1111/tops.12007.
- Groenendijk, Jeroen. 1999. The logic of interrogation. *Semantics and Linguistic Theory (SALT)* 9. 109–126. doi:10.3765/salt.v9i0.2835.
- Groenendijk, Jeroen & Martin Stokhof. 1984. *Studies in the semantics of questions and the pragmatics of answers*: University of Amsterdam PhD dissertation.
- Heim, Irene. 1982. *The semantics of definite and indefinite noun phrases*: UMass Amherst PhD dissertation.

- Heim, Irene. 1998. Anaphora and semantic interpretation: A reinterpretation of Reinhart's approach. In Uli Sauerland & Orin Percus (eds.), *The interpretive tract*, 205–246. Cambridge, MA: MITWPL.
- Jäger, Gerhard. 2012. Game theory in semantics and pragmatics. In Claudia Maienborn, Klaus von Stechow & Paul Portner (eds.), *Semantics: An international handbook of natural language meaning*, vol. 3, 2487–2425. Berlin: Mouton de Gruyter.
- Kamp, Hans. 1981. A theory of truth and discourse representation. In Jeroen Groenendijk, Theo M. V. Janssen & Martin Stockhof (eds.), *Formal methods in the study of language*, 277–322. Amsterdam: Mathematical Centre.
- Kao, Justine T., Leon Bergen & Noah D. Goodman. 2014. Formalizing the pragmatics of metaphor understanding. *Cognitive Science Society (CSS)* 36. 719–724.
- Karttunen, Lauri. 1976. Discourse referents. In James D. McCawley (ed.), *Syntax and semantics*, vol. 7: Notes from the Linguistic Underground, 363–385. New York: Academic Press. doi:10.1163/9789004368859_021.
- Koev, Todor. 2017. Quotational indefinites. *Natural Language & Linguistic Theory* 35(2). 367–396. doi:10.1007/s11049-016-9344-x.
- Lauer, Sven. 2013. *Towards a dynamic pragmatics*: Stanford University PhD dissertation.
- Martí, Luisa. 2015. The morphosemantics of spanish indefinites. *Semantics and Linguistic Theory (SALT)* 25. 576–594. doi:10.3765/salt.v25i0.3050.
- Potts, Christopher, Daniel Lassiter, Roger Levy & Michael C. Frank. 2016. Embedded implicatures as pragmatic inferences under compositional lexical uncertainty. *Journal of Semantics* 33(4). 755–802. doi:10.1093/jos/ffv012.
- Potts, Christopher & Roger Levy. 2015. Negotiating lexical uncertainty and speaker expertise with disjunction. *Berkeley Linguistics Society (BLS)* 41. 417–445. doi:10.20354/b4414110013.
- Solt, Stephanie & Jon Stevens. 2018. *Some Three Students*: Towards a unified account of *some*. *Semantics and Linguistic Theory (SALT)* 28. 345–365. doi:10.3765/salt.v28i0.4436.
- Weir, Andrew. 2012. *Some*, speaker knowledge, and subkinds. *European Summer School in Logic, Language and Information (ESSLLI), Student Session* 180–190.

Lewis Esposito
Margaret Jacks Hall
Building 460
Stanford University
Stanford, CA 94305
lesposi1@stanford.edu

Christopher Potts
Margaret Jacks Hall
Building 460
Stanford University
Stanford, CA 94305
cgpotts@stanford.edu