

An alternative semantics of the Samoan article system

M K Snigaroff

September 6, 2023

1 Introduction

Due to the difficult nature of conducting novel fieldwork for semantics and pragmatics, a great deal remains undiscovered in understudied languages around the world. Western Samoan constitutes one such case; while much has been invested in understanding the syntax of its case patterns alone (see Koopman (2012), Cook (1996), Collins (2014), and Tollan (2018), to name a few), comparatively little has been done from the perspective of any area of formal semantics. What has been done tends to rely on data from the 1996 Samoan grammar by Mosel and Hovdhaugen; for instance, Ionin (2006) mentions Samoan in her theory on the semantics of specificity in article systems, but as her data come exclusively from the 1996 grammar, it is not possible to supplement the analysis with negative data or speaker commentary. She concludes,

Based only on the few examples in Mosel and Hovdhaugen, it is impossible to develop a theory of article semantics in Samoan. However, a preliminary generalization can be made: Samoan is an example of a language which uses one article in specific environments and a different article in non-specific environments.

(Ionin 2006:218)

In this article, I endeavor to develop an in-depth semantic account of the Samoan article system, historically described as one based on specificity rather than definiteness (as mentioned here in Ionin (2006), as well as in the grammar). Using novel data, I show that there is much more than meets the eye in the system—certainly more than terms such as “specific” and “non-specific”, themselves imprecise and debated, can capture.

I develop the denotations for both articles in the system under the framework of Hamblin or alternative semantics, which originates from Hamblin’s 1973 musings on the interpretation and semantic composition of questions. Hamblin’s system takes into account the alternatives that speakers and addressees consider when interpreting an utterance (e.g., all the alternative answers to a question) by rendering all entities, properties, and propositions into sets that compose with one another. These sets consist of the alternatives that speakers consider, and are reduced over the course of the derivation into a proposition or a singleton set containing a proposition.

Hamblin semantics renders indefinite nominals such as *a dog* into sets of entities all possessing the same property (e.g., {dog₁, dog₂, dog₃}. Kratzer and Shimoyama (2002) and Kratzer (2005) take advantage of this notion to explain the behavior of the German indefinites and *ein* and *irgendein*. They suggest that *ein* can be considered an ordinary indefinite that generate a set of entities from the domain of discourse entities that have the same property. *Irgendein* appears semantically very similar to *ein*, but has a domain-widening effect on the nominal it modifies; Kratzer and Shimoyama show that its behavior can be accounted for by giving it the same denotation as *ein*, but without any domain of discourse restriction. Thus, a noun like ‘dog’ modified by *irgendein* is denoted by the set of all dogs in the world—not merely the dogs in the current discourse—resulting in a sense of domain-widening.

I suggest that the Samoan article system mimics the indefinites present in German, with the “specific” corresponding to German *ein*, and the “non-specific” corresponding to German *irgendein*. That is, the specific article generates sets of entities within the domain of discourse, while the non-specific article generates sets of all entities of the same type in the world. This results in a similar

sense of domain-widening in the interpretation of non-specific-marked nominals. The only difference I initially propose between the German indefinite system and the Samoan system, is that the specific article additionally carries a presupposition of existence, such that a nominal marked by the specific is presupposed to exist.

While this analysis accounts for a great deal of nuanced empirical facts in the distributions and interpretations of Samoan articles, it does not account for everything. One last difference between the German system and that of Samoan is that while *irgendein* can occur in virtually any environment, the non-specific is strongly dispreferred in out-of-the-blue episodic sentences—but not in episodic sentences with some level of context. Thus, I show that the importance of context cannot simply be thrown out in an analysis of the non-specific.

Although this article stops short of a full account of these discourse-related facts, it presents a sketch of a promising account. I propose that the specific and non-specific articles actually generate alternatives in quite different ways: The specific generates a set of entities *with different properties* that belong to the current discourse, while the non-specific generates a set of entities of the same property (ignoring the discourse). A higher operator later composes these sets under the there exist alternatives to the proposition generated. Because other entities are always presumed to exist (regardless of property), and therefore are present in the domain of discourse, speakers are likely to accommodate out-of-the-blue sentences containing the specific article. However, if the alternatives generated only contain entities of a certain property (as is the case with non-specific nominals), context does not necessarily provide the discourse-salient entities necessary. That is, when a specific-marked ‘dog’ is mentioned in an out-of-the-blue sentence, it is easy for speakers to imagine context in which other, alternative entities around this dog exist, as the entities can be of any kind. However, if ‘dog’ is non-specific, the alternative entities compared must be other dogs. Thus, it is more difficult for a speaker to accommodate a non-specific instance ‘dog’ out-of-the-blue, because the context requires that one dog must be compared against other dogs.

I break down my analysis as follows: § 2 gives an overview of the Samoan article system, including the syntax of articles in general, as well as the the distributions and interpretations of each article. I briefly show in this section that the specific article contains a presupposition of existence. In § 3, I lay out the basics of the Hamblin framework, and present the case study of this system from Kratzer and Shimoyama (2002) and Kratzer (2005) regarding German indefinites. In § 4, I begin my analysis of Samoan articles with a direct comparison to German articles, showing how their distributions and interpretations compare and supplementing the analysis with only the previously-established presupposition on the specific. Finally, in § 5, I reveal where precisely this analysis breaks down. I present the intuitions behind a secondary analysis, as well as some more concrete mechanics behind it, but ultimately conclude that Hamblin semantics may not be a sufficiently complex framework for this particular case.

1.1 A note on methods

The often fragile and subjective nature of semantic/pragmatic intuitions makes fieldwork in these areas rather delicate. For this reason, the selection of speakers for this project was undertaken with great care. The data below were gathered from four native speakers of Western Samoan with fluency in English. Each of the speakers were from Samoa, had grown up speaking Samoan, and had contact with the language on a day-to-day basis. Three of the four lived in Samoa at the time of elicitation (which was conducted over Zoom).

Samoan is widely known to have two registers: the vernacular (*tautala leaga*); and the literary language, often considered “formal”, off of which most grammars and linguistic studies are based (*tautala lelei*). The differences between these two are primarily lexical and phonological,¹ and they can be blurred as elements of the vernacular appear in the literary language or vice-versa (Mosel and Hovdhaugen 1992).

In my elicitations, speakers were encouraged to translate sentences as they would say them every day, but several still preferred to use the phonology of the *tautala lelei*. Thus, the data in this article

¹A subtype of written language is Biblical Samoan, whose differences are more dramatic. See Mosel and Hovdhaugen (1992:8-11) for a more detailed analysis of the varieties of register.

consist of a mix of *tautala leaga* and *tautala lelei* phonology—most notable in the collapse of the /k/-/t/ and /g/-/n/ contrasts. It is unlikely that these phonological differences had any effect on speakers’ interpretations of nominal phrases, the subject of this study.

1.2 Background on Samoan

Western Samoan (henceforth, Samoan) belongs to the “Samoic outlier” subfamily of Polynesian languages (Clemens and Massam 2021), and is spoken natively by about 500,000 people on the island of Samoa. Like other languages in the family, it is a strongly head-initial, with VSO word order and post-verbal scrambling. Samoan argument structure is somewhat controversial, but it can be descriptively labeled as an ergative-absolutive language, with ergative case overtly marked by a prefix and absolutive nominals unmarked. A third case, glossed as “locative-directional” in Mosel and Hovdhaugen (1992) and argued to be accusative in Tollan (2018), marks certain direct and indirect objects as well as representing locative case. Since the status of this case prefix (*i*) is inconsequential for my analysis, I gloss it simply as “locative”.

Samoan nominal structure reflects the general head-initial nature of the language, with case markers preceding determiners, followed by the the nucleus of the nominal.

- (1) *i le tama*
 LOC DET boy
 ‘at the boy’/‘to the boy’

Number is typically morphologically unmarked on the noun itself, although there are instances of partial reduplication to create plurals (e.g., *tama:loa* ‘man’ → *tama:loloa* ‘men’). Rather, determiners expone number.

I leave a more thorough analysis of the syntax and interpretation of Samoan nominals for the following section, in which I describe the specifics of the article system.

2 The Samoan article system

2.1 Syntax

Samoan articles can broadly be classified into two kinds², labeled by Mosel and Hovdhaugen (1992) as the “specific” and “non-specific”. The exponents of each are shown below.

	SPEC	NSPEC
sg	<i>le</i>	<i>se</i>
pl	\emptyset	<i>ni</i>

Note that the plural specific article has a null exponent; number on specific nominals is thus determined by the presence or absence of *le*.

As stated in § 1.2, Samoan articles follow case markers and precede the nouns they modify. Possessives additionally display contrastive specificity, and certain *wh*-words, can optionally as well:

- (2) a. *l-a’u fesili*
 LE-1.POSS question
 ‘my question’
 b. *s-au fesili*
 SE-2.POSS question
 ‘your question’
- (3) a. *O le a:?*
 PRES LE what
 ‘What is it?’

²Mosel and Hovdhaugen (1992) additionally present multiple categories of diminutive articles, but for the most part these can be subdivided into specific and non-specific as well. Thus, I take the primary contrast within the article system to be a binary one.

- b. O se a:
 PRES SE what
 ‘What is it?’

While the specificity morphemes appear on possessives, they do not appear on pronouns or proper names. (See § 2.2.1 for an exception to the latter.) To remain as syntactically neutral as possible, I refer to *le/Ø* and *se/ni* as “articles” throughout this article.

2.2 Distributions

While speakers commonly cite *le/Ø* as equivalent to English ‘the’ and *se/ni* to English ‘a’, the two articles possess a much more nuanced distribution—precisely what prompted Mosel and Hovdhaugen (1992) to label the system as one based on *specificity* rather than *definiteness*. However, “specificity” is unfortunately a rather unspecific term. Works such as Enç (1991), Diesing (1992), and Ionin (2006) debate whether specificity reflects partitivity, presuppositionality, or noteworthiness, among other things. Indeed, Farkas (1994) argues that there are minimally three different notions labeled as “specificity” under discussion in the sphere of theoretical semantics. For its part, the Samoan grammar, as a piece of descriptive literature, does little to clarify the precise nature of the article system dichotomy. The basic generalization arrived at is:

The specific article singular *le/l=* ART indicates that the noun phrase refers to one particular entity regardless of whether it is definite or indefinite, or to the whole class of what is denoted by the nucleus of the noun phrase. . . .

The nonspecific singular *se/s=* ART(nsp.sg) expresses the fact that the noun phrase does not refer to a particular, specified item, but to any member of the conceptual category denoted by the nucleus of the noun phrase and its adjuncts.

(from Mosel and Hovdhaugen (1992:259, 261))

As a result, novel data is required to address some of the conceptual gaps left in Samoan descriptive literature regarding the distributions of *le/Ø* and *se/ni*, as well as their semantic contributions to the interpretation of a sentence.

2.2.1 Distribution of *le/Ø*

The specific article is the default for translating most English sentences and has a wide range of possible environments. In fact, the only place that Mosel and Hovdhaugen (1992) note is illicit for *le/Ø* is sentences that assert the nonexistence of something with the verb *leai* ‘does not exist’: “A noun phrase functioning as an absolutive argument of *leai* is always non-specific” (480). That is, the absolutive argument of *leai*—its only argument—is always *se/ni*. This is true even in the case of proper names:

- (4) E leai se Seu.
 NPST not.exist SE S.
 ‘Seu is not present [There is no Seu].’

(from Mosel and Hovdhaugen (1992:263), ex. 6.62)

The ban on *le/Ø* after *leai* becomes less mysterious when one takes into account that speakers generally cannot use *le/Ø* to modify nominals which they know do not exist. This is true even when the nominals occur in questions, in conditionals, or under negation. In the below sentences, it is explicitly established between speaker and addressee that in the world, there are no queens of America or California and no giants.

- (5) *In questions:*
 a. # O ai le tupu tama’ita’i o Ameriki?
 PRES who LE king lady GEN America
 Intended: ‘Who is a queen of America?’

- (6) *In conditionals:*
- a. # E sili le fiafia o le sau'ai pe'a: alu i aikupito.
 NPST be.best LE happiness GEN **LE** giant if go LOC Egypt
 Intended: 'A giant would be happiest if he went to Egypt.'
- (7) *Under negation:*
- a. # Ou te le'i va'ai i le tupu tama'ita'i o Kalefo:nia.
 1.sg NPST NEG see LOC **LE** king lady GEN California
 Intended: 'I haven't see a queen of California.'

The sentences in (5)-(7) can be remedied if the speaker imagines himself in a world in which America and California have queens, and in which giants exist. That is, in using *le/Ø*, the speaker conveys his own knowledge that the nominal in question exists in the world. Thus, the sentences are still acceptable in certain contexts, such as speaker ignorance or intentional deceitfulness:

- (8) *Ignorance:*
- Context: You have a friend who doesn't know much about the state of the world.
 Friend:
 O ai le tupu tama'ita'i o Ameriki?
 'Who is the queen of America?'
- Response:
 E leai { se / #le } tupu tama'ita'i o Amerika!
 NPST not.exist **SE LE** king lady GEN America
 'There is no queen of America!'

- (9) *Deceit:*
- a. Na e va'ai i le sau'ai?
 PST 2.sg see LOC **LE** giant?
 'Did you see the giant?'
- Speaker comment: "All my friends used to do that. We used to do that to each other. ... We'd just stare at something and pretend we were looking at something, and there's one guy there who has no idea what's going on. [We were] like, 'Did you see that?'"
- b. Na e va'ai i se sau'ai?
 PST 2.sg see LOC **SE** giant?
 'Have you (ever) seen a giant?'

In (8), Friend believes that there is a queen of America, and thus marks *tupu tama'ita'i o Ameriki* with *le*—a felicitous utterance of (5). Note that correcting Friend's knowledge of the world requires *se*, rather than *le*. In (9a), the speaker recounts a prank he would play on his friends, in which he would pretend to see something interesting, turn to his friend, and ask if his friend had seen *le*-giant. The friend is meant to assume our speaker knows of the existence of a giant and feel disappointed for not having seen one. This sense completely disappears with *se* in (9b), a neutral question in which the speaker does not imply the existence of giants.

All this points to *le/Ø* having a presupposition of existence—that is, if a speaker felicitously uses *le/Ø* to modify a nominal, he must believe that such a nominal exists. I claim that in using *le/Ø*, he does not assert, but rather presupposes, this existence. Thus, in environments involving questions, conditionals, or negation, the existence of a *le/Ø*-marked nominal cannot be denied; rather, as a presupposition, it cannot be canceled. I offer this brief assessment of *le/Ø*'s distribution for now, and will formalize the notions here in § 4.

Note that another restriction on the distribution of *le/Ø* also exists: it cannot occur after the quantifier *so'o* 'every, each'. Instead, *se* must occur.

- (10) So'o { *le / se } aiga lava e iai se tagata pisa.
 every **LE SE** family EMPH NPST be.there SE person noisy
 'In every family there is a noisy person.'

I refer to *so'o* as a “quantifier” here purely because it seems to have some kind of quantificational force. However, its actual type and denotation are may be more controversial; in her analysis of other quantifier-like items such as *uma* ‘all’, Hohaus (2018) proposes that Samoan “quantifiers” are rather “pseudo-quantifiers”. A complete analysis of the semantics of *le/Ø* and *se/ni* that involves this small detail would require a thorough semantics of *so'o* itself; given these difficulties, I choose to leave this issue for future work.

2.2.2 Distribution of *se/ni*

The distribution of *se/ni* is more nuanced than that of its counterpart. Speakers generally find episodic sentences containing *se/ni* questionable, amending them most commonly by giving the sentence polar-question intonation or modifying the relevant nominal with a relative clause.

(11) *Making the sentence interrogative:*

- a. *E to'atolu fafine e nonofo i se nofoa.
 NPST two woman NPST sit.pl LOC **SE** chair
 Intended: ‘The two women sit on a chair.’
- b. E to'atolu fafine e nonofo i se nofoa?
 NPST two woman NPST sit.pl LOC **SE** chair
 ‘Are two women sitting on a chair?’

(12) *Adding a relative clause:*

- a. ?? Sa fafaga e le tama se maile.
 PST feed ERG LE boy **SE** dog
 Intended: ‘The boy fed a dog.’
- b. Sa fafaga e le tama se maile Chihuahua.
 PST feed ERG LE boy **SE** dog Chihuahua
 Intended: ‘The boy fed some Chihuahua dog.’

As Giannakidou (2011) states, being illicit in an episodic sentence is “the hallmark property” of negative polarity items (NPIs). Indeed, the environments in (11)-(12) are all possible environments for, e.g., the English *any*, a classic polarity-sensitive item. Some more polarity-related environments which allow *se/ni* include:

(13) *Under negation:*

- a. E le'i tusia e Tai se upu!
 NPST NEG write ERG T. **SE** word
 ‘Tai didn’t write a [single] word!’

(14) *In imperatives:*

- a. Piki se pepa.
 pick **SE** card
 ‘Pick a card.’

(15) *After attitude verbs of uncertainty:*

- a. Ailuga o se maile ga 'aia le i'a.
 be.doubtful PRES **SE** dog PST eat LE fish
 ‘I doubt a dog ate the fish.’

(16) *In the antecedent of a conditional:*

- a. Afai ae kaukala leokele se kamaikiki aoga, o le'a: alu ese Kai mai le falekusi.
 if or talk loud **SE** child school FUT go out K. away.from LE library
 ‘If a student talks loud, Kai will leave the library.’

Besides in the context of episodic sentences, *se/ni* is also dispreferred in the consequents of conditionals, as shown in (17a)—another typical behavior of NPIs. However, sentences like (17a) can

curiously be remedied simply by substituting the TAM marker *e* for *o le'a*: (which I have glossed as ‘NPST’ and ‘FUT’, respectively). Speakers comment that (17b) sounds more acceptable because *o le'a*: “sounds more uncertain” than its TAM counterpart *e*.

(17) *Consequent of a conditional*:

- a. ? A alu ese Tai ma le fale e tu: **se** manulele lona 'ulu
 if go out T. and LE house NPST land **SE** bird 3.sg.POSS head
 Intended: ‘If Tai goes out of his house, a bird will land on his head.’
 Speaker comment: “*Definitely* a bird will land on his head.”
- b. A alu ese Tai ma le fale o le'a: tu: **se** manulele lona 'ulu
 if go out T. and LE house FUT land **SE** bird 3.sg.POSS head
 ‘If Tai goes out of his house, a bird will land on his head.’
 Speaker comment: “There’s a *probability* of a bird landing on his head. It’s gonna happen, but not quite sure...”

Similarly, *se/ni* is dispreferred after the future-marker *la'a*, with speaker citing the “certainty” of *la'a* as clashing with *se*. I gloss *la'a* below as “realis future” to distinguish it from *o le'a*: ‘FUT’.

- (18) La'a 'ai e { **le** / ??**se** } tama le niu.
 FUT.REAL eat ERG **LE SE** boy LE coconut
 ‘The boy will eat the coconut.’
 Speaker comment: “*La'a* sounds too certain [with *se*]. . . sounds like you are predicting the future.”

Perhaps more perplexingly: While I have specifically labeled attitude verbs of uncertainty as licit environments for *se/ni*-marked nominals (see (15)), it is also possible for *se/ni* to exist under any sort of attitude verb—and even more generally in other embedded contexts such as reportative verbs.

(19) *After attitude verbs of certainty*:

- a. E kalikogu Kai e mafai e **se** kagaka poko oga fo:'ia loga fa'afikauli.
 NPST believe K. NPST be.able ERG **SE** person wise that resolve 3.sg.POSS problem
 ‘Kai believes there is a wise person who can solve his problem.’
- b. E iloa lelei e Simi na o:mai **ni** maile i totonu o le fale.
 NPST know well ERG S. PST come.pl **SE** dog LOC inside GEN LE house
 ‘Simi knows well that dogs came into the house.’

(20) *After reportative verbs*:

- a. Na musumusu Simi na 'ai e le pusi **se** isumu.
 PST whisper S. PST eat ERG LE cat **SE** mouse
 ‘Simi whispered that the cat ate a mouse.’
- b. Na vala'au le tama'ita'i ua gaoi e **se** tama:loa lana atou.
 PST shout LE lady PST.PERF steal ERG **SE** man 3sg.POSS bag
 ‘A lady shouted that a man stole her bag.’

The sentences in (19)-(20) do not contain typical licit environments of other NPIs, or even more generally of polarity-sensitive items. Thus, the distribution of *se/ni* appears more complicated than that of a canonical NPI.

2.3 Interpretations

2.3.1 The contrast is not referentiality or scope

Since the environments of *le/Ø* and *se/ni* overlap, why might a speaker choose one over the other? When speakers are given minimal pair sentences with *le/Ø* and *se/ni*, they often characterize the difference as one of referentiality or scope, in which *le/Ø* might be considered the referential or wide-scope counterpart to *se/ni*.

- (21) a. Ou te le'i faitaua lava **se** tusi.
 1.sg NPST NEG read EMPH **SE** book
 'I have not ever read a book.'
 Speaker comment: "Maybe you can't read."
 b. Ou te le'i faitaua lava **le** tusi.
 1.sg NPST NEG read EMPH **LE** book
 'I have not read the book.'
 Speaker comment: "You had assigned reading [i.e., you had to read a particular book]."
- (22) a. Ou te iloa o **se** maile na aia le i'a.
 1.sg NPST know PRES **SE** dog PST eat LE fish
 'I know a dog ate the fish.'
 Speaker comment: "You don't know which dog."
 b. Ou te iloa o **le** maile na aia le i'a.
 'I know the dog ate the fish.'
 Speaker comment: "*O le* is 'the'. So, some specific dog ate the fish."

In the above sentences, speakers were given minimal pairs one after the other and asked about interpretational difference between the two. They tended to describe *le/∅* as denoting a "specific" individual, while *se/ni* nominals did not do so. Examples like (21) and (22) point to referentiality or scope as the semantic distinction between the two articles. For instance, perhaps *le* scopes over negation in (21b), while *se* is unable to do so in (21a); or, perhaps *le* is simply inherently referential, while *se* is not. Likewise, we might interpret *le* in (22b) as scoping over the attitude verb *iloa* 'know' or making reference to a particular dog.

However, further prodding reveals that referentiality and wide scope are not inherent properties of *le/∅*. For instance, the *le*-sentence in (22b), in which the speaker described the fish-eating culprit as "some specific dog", can be licitly followed by:

- (23) Ou te iloa o **le** maile na aia le i'a. ...

Ae ou te le: iloa po'o ai!
 but 1.sg NPST NEG know which one

'I know a dog ate the fish. . .but I don't know which one!'

In (23), the *le*-marked nominal is clearly non-referential, since the speaker is unaware of the identity of the dog. After describing *le maile* and other nominals as referring to particular entities, the Samoan language consultants themselves were sometimes mildly surprised that follow-ups such as the one in (23) were possible. However, when they did not compare *le*-marked nominals to *se*-marked minimal pairs, speakers had no trouble identifying the non-referentiality of *le/∅*.

- (24) E talitogi Tai e mafai e **le** tagata poto ona fofo: lona fa'afitauli.
 NPST believe T. NPST be.able ERG **LE** person wise that cure 3.sg.POSS problem
 'Tai believes that there is a wise person who can solve his problem.'
 Speaker comment: "In this particular case, it's *any* wise person. There *is* a wise person that can solve this problem. It's either a particular person, or he believes there *is* a wise person."

The speaker above also offered a clearly narrow-scope instance of *le/∅* :

- (25) So'o se aiga lava **le tagata pisa**.
 every SE family EMPH **LE person noisy**
 'In every family there's a noisy one [person].'

The sentence in (25) does not mean that each family possesses the same noisy person; rather, *le tagata pisa* 'a noisy person' takes narrow scope with respect to *so'o* 'every'. The speaker reported that substituting *le* for *se* here resulted in exactly the same meaning.

2.3.2 Uncertainty

A more reliable interpretational contrast between *se/ni* and *le/Ø* is that the former adds some degree of uncertainty while the latter does not. The nature of this uncertainty can vary; speakers may characterize *se/ni* -sentence as conveying uncertainty about which entity took part in an event, or uncertainty about whether an event took place at all. The sentences in (26) display the certainty contrast between *le/Ø* and *se/ni*, with the two speaker comments in (26b) displaying the contrast in “uncertainty” interpretations.

For instance, even when *se/ni* appears after an attitude verb like *iloa* ‘know’, speakers report that the sentence imbues a sense of uncertainty. With *le/Ø*, this is not the case.

- (26) a. Ou te iloa o **le** maile na aia le i’a.
 1.sg NPST know PRES **LE** dog PST eat LE fish
 ‘I know a dog ate the fish.’
- b. Ou te iloa o **se** maile na aia le i’a.
 ‘I know a dog ate the fish.’
 Speaker A comment: “You’re like, 90% sure.”
 Speaker B comment: “Maybe you’re not sure which dog ate the fish.”

Even though (26a) and (26b) both translate to ‘I know a dog ate the fish’, they carry a subtle semantic difference. Sentence (26a) straightforwardly conveys that the speaker knows the event took place, and that a dog was involved. However, the interpretation of (26b) appears more nuanced, with Speaker A questioning whether the event in question occurred, and Speaker B expressing uncertainty about who exactly took part in it. Note that even though the matrix attitude verb is *iloa* ‘know’, uncertainty about event occurrence is still possible.

Follow-up statements to sentences like those in (26) can add doubt to the identity of participants in both *le/Ø* - and *se/ni* -sentences. In (27), the speaker states that Simi has incomplete knowledge about the agent of an event, following up the assertion with, “but he doesn’t know which ones [came into the house]!” Though both the *le/Ø* and *se/ni* variants of the sentences appear to have the same truth values, speaker commentary reveals that there remains a subtle difference between the two.

- (27) a. E iloa lelei e Simi na o:mai **Ø** maile i totonu o le fale ...
 NPST know well ERG S. PST come LE.pl dog LOC inside GEN LE house
 ae le: mautinoa po’o ai!
 but he doesn’t know which ones
 ‘Simi knows well that dogs came into the house...but he doesn’t know which ones!’
- Speaker comment: “Could still be that they’re his dogs, or some other dogs; the neighbor’s dogs, or his dogs bringing their friends.”
- b. E iloa lelei e Simi na o: mai **ni** maile i totonu o le fale...ae le: mautinoa po’o ai!
 ‘Simi knows well that dogs came into the house...but he doesn’t know which ones!’
- Speaker comment: “It implies that, maybe he doesn’t have dogs! So we’re not quite sure where the dog prints came from.”

In (27a), Simi considers a set of dogs he knows of, despite not being aware of the particular identities of the dogs that came into the house. In (27b), Simi expresses a greater degree of bafflement at the identities of the dogs—he doesn’t have dogs, so there are no contextually-relevant individuals he might ascribe the dog prints to. Simi is more perplexed in (27a) than in (27b), reflecting a greater degree of uncertainty about the *se/ni* -marked nominal than the *le/Ø* -marked nominal.

In summary, the Samoan articles *le/Ø* and *se/ni* display very subtle differences in distribution as well as interpretation. While the former can occur almost anywhere, save in assertions of nonexistence and in the scope of the quantifier *so’o* ‘every, each’, the latter is unacceptable in out-of-the-blue episodic sentences. Semantically, the contrast between *le/Ø* and *se/ni* cannot be described as one of referentiality or scope; both can be referential or nonreferential, taking wide or narrow scope. The interpretations of each article rather pertain to some notion of uncertainty: *se/ni* -marked nominals

may be interpreted as conveying doubt about the occurrence of an event, or about the identities of the event’s participants, while *le/Ø*-marked nominals carry no such meaning.

In the following section, I lay out the framework necessary to capture a semantic analysis, as well as a helpful case study within said framework that exhibits facts reminiscent of the ones here.

3 An alternative semantics of indefinites

Innumerable theories exist to account for the behavior of indefinite articles. I will be adopting a framework originally proposed by Hamblin (1973), which itself adjusts Montague (1970)’s framework of English semantics. Hamblin argues that Montague grammar should not ignore the semantics of interrogatives, but rather already has the basic machinery to account for them. Rather than treating all compositions as outputting individuals, properties, or propositions, Hamblin proposes that semantic elements compose to form sets with members of these types. This allows questions, which are clearly non-propositional, to be sets of propositions—the members of the set being all possible answers to the question. This basic intuition about the meanings of indeterminate expressions has been expanded upon in later works (e.g., Kratzer and Shimoyama 2002, Kratzer 2005). Below, I sketch out Hamblin (or “alternative”) semantics, paying special attention to Kratzer and Shimoyama (2002)’s and Kratzer (2005)’s particular version, since they have applied the analysis to indefinites.

3.1 Hamblin semantics

Kratzer and Shimoyama (2002) summarize the framework as one in which “indeterminate phrases . . . introduce sets of alternatives that keep ‘expanding’ until they meet an operator that selects them” (4). These sets of alternatives might consist of propositions, individuals, or properties, and as multi-membered sets, they are uninterpretable. However, operators such as negation, the existential operator, the universal operator, questions, etc. can take these sets and output sets of propositions or properties, ultimately rendering a singleton set of one proposition.

Hamblin’s original proposal used sets to account for questions, pointing out that while *wh*-words such as *who* or *what* undeniably feel “noun-like” (and *where* “adverb-like”, etc.), they cannot be said to belong to the class of individuals as proper and definite nouns do. Rather, Hamblin suggests they represent sets of individuals, where *who* is the set of all humans, *what* the set of all objects, etc. However, in such a semantics it cannot be the case that only *wh*-words constitute sets. Rather, all nominals can be represented this way, with definites such as *Mary* denoting a set consisting of one individual: *Mary*.

Because nominals now constitute sets of individuals, they cannot compose via Function Application, or other methods of composition involving non-sets. Properties can be taken to constitute singleton sets themselves, the only member of the being the function denoting that property. They compose with nominals (sets of individuals) through a pointwise function application rule, e.g.:

(28) Hamblin Function Application

If α is a branching node with daughters β and γ , where $[[\beta]]^{w,g} \subseteq D_\sigma$ and $[[\gamma]]^{w,g} \subseteq D_{\sigma\tau}$, then:
 $[[\alpha]]^{w,g} = \{ \alpha \in D_\tau \mid \exists b \exists c [b \in [[\beta]]^{w,g} \ \& \ c \in [[\gamma]]^{w,g} \ \& \ a = c(b)] \}$

The rule above simply applies the function contained in a singleton set to each member of a set of individuals. Its application might result in a derivation like the following, in this case of *who slept*:

(29) *who* denotes the set individuals containing all humans:
 $\{ \text{person}_a, \text{person}_b, \text{person}_c, \dots \}$

(30) *slept* denotes a singleton set whose member is the property ‘slept’:
 $\{ \lambda x. \lambda w'. \text{slept}'(x)(w') \}$

(31) *who slept* denotes the set of propositions:
 $\{ \text{person}_a \text{ slept}, \text{person}_b \text{ slept}, \text{person}_c \text{ slept}, \dots \}$

(adapted from Kratzer and Shimoyama (2002:6))

In the above, (29) and (30) compose via pointwise function application to form (31), such that ‘slept’ applies to each of the members of the set in (29). The question *Who slept?* might encounter a Q-operator such as the following:

$$(32) \quad [[Q\alpha]]^{w,g} = \{ \lambda w'. \forall p [p \in [[\alpha]]^{w,g} \rightarrow [p(w) = 1 \leftrightarrow p(w') = 1]] \}$$

In (32), α is the set of all propositions (e.g., those in (31)), and the Q-operator itself outputs a singleton set: the answer to *who slept* in world w .³ However, one can find a simpler example in the operators that compute indefinites, such as the existential operator below.

$$(33) \quad [[\exists\alpha]]^{w,g} = \{ \lambda w'. \exists p [p \in [[\alpha]]^{w,g} \ \& \ p(w') = 1] \}$$

Once again, α is the set of propositions; if we take the sentence *A man slept*, these might be $\{ \text{man}_a \text{ slept, man}_b \text{ slept, man}_c \text{ slept, } \dots \}$, and so on for all men. (33), then, reduces the potentially infinite set into a singleton containing the proposition that there exists a proposition that is true in the world and that can be found among the set of all men sleeping. In other words, that *a man slept*.

Besides Q-operators and the existential operator, there may also exist operators such as negation and the universal operator, as well as operators over sets of individuals and any other logically possible kind of operator. Some of these operators may be overt in certain languages (e.g., Japanese *ka* ‘Q’ in Kratzer and Shimoyama (2002)) and unpronounced in others.

There are multiple reasons one might adopt a Hamblin-style semantics over some other framework. Falauş (2013) names questions (already discussed here), implicatures, and focus as three areas which have particularly benefited from the framework. Hamblin semantics allows one to formalize the basic Gricean intuitions behind implicature: that is, that an addressee may infer the falsity of a stronger statement that could have been uttered, because the speaker intentionally chose to assert a weaker statement. As Falauş (2013) explains, Hamblin semantics actually posits mechanics to explain how addressee generates the set of possible alternatives to the utterance that was made (9). To account for focus, the Hamblin framework has been expanded by Rooth (1992) into a “two-dimensional” alternative semantics, in which alternatives are compositionally built up in parallel to the primary semantic meaning of an utterance. These alternatives represent the “focus value” of the utterance—that is, the set of alternatives against which an utterance is focused.

Alternative semantics has also proven useful in explaining indefinites and NPIs, two topics which are particularly relevant here. Under Hamblin’s analysis, indefinites, like *wh*-words, will correspond to sets of individuals. This explains an otherwise coincidental typological fact: It is crosslinguistically common for a language to have its indefinites and *wh*-words be the same, or etymologically related. As for NPIs, Krifka (1995) shows how these fit naturally into a Hamblin semantics system, as they are typically thought to involve some sort of domain-widening and thereby consideration of a wider domain of alternatives. He argues that in considering NPIs to be “just a special case of other constructions that introduce alternatives” (255), suggesting that manipulating the sorts of alternatives they generate can help in deriving at least some of the diversity of NPIs.

3.2 Case study: *Irgendein*

The Hamblin system makes the important prediction that indefinites themselves do not have any existential force, but are rather interpreted as existential due to an unpronounced existential operator. Kratzer and Shimoyama (2002) and Kratzer (2005) illustrate the point with Japanese *wh*-words and the German *irgend-* series (which I henceforth refer to as *irgendein*). Here, I will focus on *irgendein*, since it will become apparent that it shares many similarities with Samoan *se/ni*.

Irgendein is similar to in that “it displays what we might describe as ‘polarity sensitivity’” (Kratzer and Shimoyama 2002) (in terms of its interpretation), but displays a more flexible distribution than typical polarity-sensitive items. The examples below were taken from Chierchia (2013).

$$(34) \quad \text{Niemand hat } \mathbf{irgend}\text{-ein} \text{ Buch mitgebracht.} \\ \text{no one} \quad \text{has } \mathbf{IRGEND}\text{-a book brought}$$

³This particular Q-operator comes from Groenendijk and Stokhof (1984). However, it may be the case that a Q-operator outputs all possible answers, or all true answers.

‘No one has brought along any book.’

(from pg. 255, ex. 20a)

- (35) Nehme **irgend**-ein Karte.
take **IRGEND**-a card

‘Take a [any] card.’

(from pg. 249, ex. 7c)

- (36) Du darfst **irgend**-eine Kuche essen.
2p may **IRGEND**-a cake eat

‘You may eat any cake.’

(from pg. 249, ex. 6b)

- (37) **Irgend**-ein Student hat angerufen.
IRGEND-a student has called

‘Some student has called.’ (for all the speaker knows, it might be any student)

(from pg. 256, ex. 22a)

- (38) John ist hinausgelaufen un hat an **irgend**-eine Tuer geklopft
J. is run.out and has on **IRGEND**-a door knocked

‘John ran out and knocked at a door.’ (for all John cared, he might have knocked at any door)

(from pg. 257, ex. 22b; glossing courtesy of Darragh Winkelman, personal communication)

In (34) and (35), *irgendein* appears in contexts associated with polarity-sensitive items: under the scope of a negative quantifier⁴ (in (34)), in an imperative (in (35)), and after a modal (in (36)). In (37)-(38), *irgendein* appears in clearly non-polarity-related contexts—specifically, episodic sentences. Here, the meaning of *irgendein*, rather than being that of a “plain” indefinite, carries a sense of speaker/agent ignorance (as in (37)) or indifference (as in (38)).

I (and Chierchia (2013), along with many others) gloss *irgendein* as bimorphemic—compositionally related to the German plain indefinite *ein*. In the analysis of Kratzer and Shimoyama, this is no accident. Their denotations for *ein* and *irgend-* are given below for all possible variable assignments g and worlds w . Importantly, the domain variable D , representing the set of possible individuals, can be valued by context.

- (39) a. $g(D) \subseteq D$

b. For *ein Mann* ‘a man’:

$$[[\text{ein}_D \text{ Mann}]]^{w,g} = \{ x \mid x \text{ is a man in } w \ \& \ x \in g(D) \}$$

- (40) For $[[\alpha]]^{w,g} \subseteq D_e$:

a. $[[\text{irgend-}\alpha]]^{w,g} = \{ x \mid \exists g'[x \in [[\alpha]]^{w,g'}] \}$

b. For *irgendein Mann* ‘IRGEND-a man’:

$$[[\text{irgend-} [\text{ein}_D \text{ Mann}]]]^{w,g} = \{ x \mid \exists g'[x \text{ is a man in } w \ \& \ x \in g'(D)] \}$$
$$= \{ x \mid x \text{ is a man in } w \}$$

The difference between $[[\text{irgend-ein Mann}]]^{w,g}$ and $[[\text{ein Mann}]]^{w,g}$ is small, but reflects the equally subtle interpretational difference between *ein Mann* and *irgendein Mann*: while the truth values of sentences containing either nominal may be the same, *irgendein Mann* forces the addressee to consider a larger set of men than *ein Mann* due to *irgend-* inducing domain-widening. That is, in (39b), the addressee considers the set of contextually-relevant men in the world, while in (40b), the addressee must consider all possible men in the world.

Both *ein*-nominals and *irgendein*-nominals yield a set of individuals, rather than an element of type e or et , so they must later be subject to the kinds of operators previously described. Kratzer and

⁴Technically speaking, *irgendein* cannot appear under sentential negation. Kratzer and Shimoyama (2002) explain this through the particulars of their analysis: *irgendein* and other indefinites “associate with” uninterpretable features such as [Neg], [\exists], [\forall], etc. and must be bound by the particular (unpronounced) operator they correspond to. According to Kratzer and Shimoyama, *irgendein* can only possess the uninterpretable feature \exists , so it cannot be closed by a [Neg] operator. However, it is just as well if an indefinite possesses any combination of uninterpretable features, so this fact is inconsequential for the present analysis of *se/ni*.

Shimoyama (2002), who are primarily interested in *irgendein*'s interaction with modals, suggest that all modals inherently carry an existential operator \exists ; this would explain *irgendein*'s presence in (36), where it occurs in the scope of a modal. However, *ein* and *irgendein* occur in a much larger swathe of environments, so some sort of operator must render them interpretable elsewhere too.

Kratzer and Shimoyama suggest that operators like \exists , \forall , \neg , or Q might be carried by certain inflectional aspects, negation, or *wh*-complementizers, among other things. This analysis means that the existential “feel” of *irgendein* (and *ein*) does not fall out from its own semantics, but rather from the higher operators that can close it. Kratzer and Shimoyama’s list of potential operators place each one in the syntax, but as we will see, this may not always be the case.

4 A pure Kratzer and Shimoyama (2002) analysis

I analyze the semantics of the Samoan article system under a framework of Hamblin semantics and demonstrate how the *le/Ø* / *se/ni* dichotomy may not be so different from the German *ein/irgendein* dichotomy, as described by Kratzer and Shimoyama (2002). The analysis must minimally account for the two sorts of uncertainty expressed in *se/ni* -sentences, as well as the distribution of *se/ni*. (Recall that the distribution of *le/Ø* is explained with its presupposition of existence, as described in 2.2.1.)

4.1 Uncertainty about existence/occurrence

As I alluded to previously, *le/Ø* might be compared to the plain German indefinite *ein*, an existential with similarly wide distribution and “neutral” meaning whose denotation as proposed by Kratzer and Shimoyama is the simplest possible denotation for an indefinite. The denotation outputs a set of all the contextually-relevant men in the world (see (39)).

However, it cannot be that *le/Ø* is truly as “plain” an indefinite as *ein*, because its distribution is slightly more restricted. As discussed in § 2.2.1, *le/Ø* possesses a presupposition of existence—a very slight deviation from *ein*. The denotation for *le/Ø* might therefore be something like (41), with a sample derivation in (42).

- (41) For domain variable δ , whose value is fixed by the assignment function g such that $g(\delta) \subset D$, where D is the domain of discourse:
- a. p is a property ($p \subseteq D_{\sigma\tau}$)
 - b. $[[le_{\delta,i} p]]^{w,g}$ is defined only if $\exists x$ in w such that $p(x) = 1$.
When defined,
 $= \{ x \mid p(x) = 1 \text{ in } w \ \& \ x \in g(\delta) \}$
- (42) For *le tama:loa* ‘LE man’:
- a. $[[le_{\delta,i} \text{tama:loa}]]^{w,g} = \{ x \mid x \text{ is a } \text{tama:loa} \text{ in } w \ \& \ x \in g(\delta) \}$
 $= \{ \text{tama:loa}_1, \text{tama:loa}_2, \text{tama:loa}_3, \dots \}$

(41) is a restatement of the definition of *ein*, but with an added condition on when *le p* can be interpreted. (42) shows the derivation of *le tama:loa*, which results in a set of individuals whose members are all the discourse-relevant men (that is, men in $g(\delta)$).

Uncertainty about event occurrence present in *se/ni* sentences can be explained via this presupposition on *le/Ø* as a result of compared alternatives. Using *le/Ø* presupposes speaker knowledge regarding the real-world existence of the nominal in question. Thus, whenever a speaker uses *se/ni*, he has consciously made the decision to not invoke this presupposition. Under classic Gricean reasoning, or a Maximize Presupposition rule of the sort in Heim (1991), an addressee can interpret this as speaker uncertainty: The speaker has made the decision to not imply his knowledge of the existence of one of the participants in the event. If one of the participants may not exist, then the occurrence of the event itself is called into question.

The below consultant commentary is particularly enlightening:

- (43) Sa fai mai Simi na 'ai e se maile le i'a.
PST say to.me S. PST eat ERG **SE dog** LE fish

‘Simi told me a dog ate the fish.’

Speaker comment: “He’s not sure. . . otherwise he would have said *le*.”

The consultant in the above sentence not only shows awareness of the alternatives to the article in question, but also how the alternatives affect the meaning of the sentence. Because Simi chose to report only that *se*-dog ate a fish, rather than the more informative *le*-dog, it follows from Gricean reasoning that he cannot make the stronger claim that such a dog necessarily exists. The result is a sense of uncertainty on the part of Simi, and in general with *se/ni* -sentences.

4.2 Domain-widening and uncertainty about identity

Following the analogy to the *ein/irgendein* dichotomy, I define *se/ni* along the lines of *irgendein* in Kratzer and Shimoyama’s article.

- (44) a. p is a property ($p \subseteq D_{\sigma\tau}$)
b. $[[se\ p]]^{w,g} = \{x : p(x) = 1 \text{ in } w\}$
- (45) For *se tama:loa* ‘SE man’:
a. $[[se\ tama:loa]]^{w,g} = \{x : x \text{ is a } tama:loa \text{ in } w\}$

While *le tama:loa* denotes the set of contextually-relevant men, *se tama:loa* picks out all men possible in the world w irrespective of the contextually-driven domain D . (Compare to (41) in the previous section.) The result is a sense of domain-widening.⁵ Note that while the denotation of *irgend-* composes with that of *ein* to produce domain-widening, *se* (and *ni*) does not appear to be morphologically complex like *irgendein*, and its denotation can be simpler.

Is this domain-widening attested in speaker interpretations of *se/ni*? Certainly, *se/ni* carries the same flavor of speaker indifference or ignorance that *irgendein* does. Domain-widening could additionally account for speakers’ initial judgments that *le/Ø* was wide scope or referential, and their general tendency to translate *le/Ø* as English ‘the’, referring to a “specific” entity, and *se/ni* as ‘a’. That is, *se/ni* is only different from *le/Ø* in that the domain of the alternatives considered with the former is larger; in *le tama:loa* ‘le man’, one must only consider other contextually-relevant men (if they exist) as alternatives, whereas with *se tama:loa* ‘se man’, one considers all other men in the world as alternatives. Because the domain of alternatives of a *le/Ø* -marked nominal is more restricted, speakers might interpret this as a higher degree of specificity.

More compelling evidence exists that *se/ni* has a domain-widening effect. For instance, the following sentence with *se* only sounds appropriate with a particular domain-widening context:

- (46) Context: Mikaele’s window is broken and. . .
- a. . . he is an old man and thinks one of the kids in the neighborhood did it.
- Fai mai Mikaele na ta’e le fa’amalama i **se tamaititi**.
say to.me M. PST break LE window by **SE child**
- ‘Mikaele told me the window was broken by **a kid**.’
- b. . . he thinks one of his three children—Soi, Mani, or Malama—did it.
*Fai mai Mikaele na ta’e le fa’amalama i **se tamaititi**.

The sentences in (46) are minimal pairs, differing only in the size of the domain of alternatives considered for *se tamaititi* ‘se kid’. In (46b), the speaker considers a more restricted set of alternatives—his three children—than in (46a), where he considers any kid around. The result is that the nominal in question in (46b) is incompatible with *se*, while in (46a), the domain of alternatives considered is sufficiently large for *se* to be used.

Domain-widening can explain the second aspect of the “uncertainty” previously alluded to in *se/ni* -sentences. Recall that some speakers indicated that *se/ni* -marked nominals created an uncertainty as to whether an event had taken place—that is, whether the nominal in question actually existed. This

⁵Technically, no preexisting domain has been widened, and thus “domain-widening” might be considered an imprecise term. However, I choose to employ it here because *se/ni* ’s extremely large domain will naturally be compared by speakers to *le/Ø* ’s comparatively narrow one; thus, the same effect may be achieved.

was explained by the presence of the alternative le/\emptyset , which carries a presupposition of existence. At the same time, some speakers expressed the uncertainty present in rather different terms:

- (47) Ou ke iloa o se maile ga 'aia le i'a.
 1.sg NPST know PRES **SE** dog PST eat LE fish
 'I know a dog ate the fish.'

Speaker: "*O se* is an unsure way of saying it. You don't know *which* dog."

MKS: "With this scenario, maybe there are a whole bunch—maybe I have, like, ten dogs. . ."

Speaker: "Yes, yes. And you don't know which one ate it."

- (48) Ou te iloa o se maile na aia le i'a.
 'I know a dog ate the fish.'

Speaker A comment: "You're like, 90% sure."

Speaker B comment: "Maybe you're not sure which dog ate the fish."

(reproduced from (26b))

In (47), the speaker indicates uncertainty not about the existence of a dog, but rather which dog among a domain of alternatives was the agent. In (48), both kinds of uncertainty are expressed: Speaker A is unsure if the event took place, while Speaker B is once again unsure of who among a set of dogs ate the fish. Speaker A may be considering the alternative le/\emptyset , in which the agent dog must exist; Speaker B, meanwhile, could be considering the complete domain of alternative dogs in w generated by se/ni . The extremely wide domain implies that Speaker A is not committed to a particular dog, or even to any of the dogs in the relevant contextual domain. Speaker B interprets this as uncertainty as to which dog of the domain committed the act, once again via classic Gricean reasoning.

The presupposition on le/\emptyset and the domain-widening of se/ni thus explain a number of interpretational effects, particularly the uncertainty present in se/ni . I have so far defined this domain-widening as the same sort specifically put forth by Kratzer and Shimoyama (2002) for *irgendein*, which has its advantages— se/ni and *irgendein* have similar "indifference" and "uncertainty" effects, and their distributions are similarly broad yet NPI-like. However, it cannot be the case that se/ni and *irgendein* are semantically identical, because their distributions are actually not the same: while *irgendein* can occur in just about any context, se/ni sounds unacceptable (or at the very least, questionable) in out-of-the-blue episodic sentences. There is nothing in the definition of se as it stands in (44) that should prohibit this. Thus, it appears there is something more going on with the semantics of se/ni that has not yet been accounted for.

5 Alternatives of different kinds

5.1 The puzzle of episodic sentences

As stated previously, resistance to episodic sentences, is one of the unifying properties of NPIs or polarity-sensitive items, and a property which se/ni at first glance seems to straightforwardly possess (in contrast with *irgendein*). Below I recapitulate some examples illustrating this characteristic.⁶

- (49) ??Sa fafaga e tama:loloa se malie.
 PST feed ERG man.pl **SE** shark
 Intended: 'The men fed a shark.'

- (50) *Sa siva se tama'ita'i i le a'oauli.
 PST dance **SE** lady LOC LE afternoon
 Intended: 'A lady danced in the afternoon.'

⁶Sentences (49)-(50) involve a single action in the simple past, the clearest instantiation of episodicity, but any sentence referencing a specific event might be considered episodic (Carlson 2005).

However, the facts are slightly more complicated. While Samoan speakers are consistently hesitant to accept out-of-the-blue episodic sentences like those in (49)-(50), they accept and even naturalistically produce episodic sentences with *se/ni* provided there is sufficient context for the sentence.

- (51) Sa fafaga e tama:loa se malie.
 PST feed ERG man SE shark

The men fed a shark.

Context offered: “A Samoan comes along and says, ‘My shark wasn’t fed!’ Hmm...the men said they fed a shark...”

- (52) a. Sa ou ui atu i le tumu o tagata.
 ‘I was walking through a big crowd.’
 b. Ae te’i, ua tamo’e mai se tama:loa ia: te a’u.
 but suddenly PST.PERF run to.1p SE man to me
 ‘But suddenly, a man ran up to me.’
 c. Fai mai tama:loa ua ou ma:lo: i le miliona tala!
 ‘The man told me I won a million Tala[Samoan currency]!’

The sentence in (51) is a minimal pair with that in (49), but is now considered acceptable by the speaker. I constructed the sentence, and the speaker offered a highly specific context to accommodate it: A Samoan complains that his shark wasn’t fed, but others respond with (51), claiming that *some* shark was fed—although it is uncertain which shark, and implied that perhaps the wrong shark was fed. The same elements of uncertainty about *which* shark are present, as discussed in § 4.2. Additionally, the context in (51) assumes the existence of multiple alternative sharks which might have been fed.

Meanwhile, the *se*-containing episodic sentence in (52b) is more naturalistically produced by a speaker as he constructs a brief narrative. The context, taking the form of the narrative itself, is explicitly provided in Samoan rather than being imagined and conveyed in English by the speaker. Once again, the context presumes the existence of multiple people in a large crowd, from which one particular person emerges as the subject of (52b).

These data are problematic for the current analysis because they can neither be explained by a direct comparison with *irgendein*, nor by any general properties of polarity-sensitive items. If the former were the case, we would expect (49)-(50) to be acceptable at all times, regardless of context; however, in the latter scenario, (51) and (52) should be unacceptable despite the salience of context. Thus, the task that remains is to reconcile the current Hamblin-based system so far developed for Samoan articles with these more nuanced facts.

5.2 Intuitions about different kinds of alternatives

The definitions set forth for *le/Ø* and *se/ni* involve the Kratzer and Shimoyama (2002) definitions for *ein* and *irgendein*, supplemented only by a presupposition on *le/Ø* to derive the environment of *le/Ø* and certain types of uncertainty in the interpretation of *se/ni*. I argue that this alone is insufficient to derive the contrast between the two articles. Currently, the alternatives generated by *le/Ø* and *se/ni* differ only in the degree to which context—that is, the domain set $g(\delta)$ of entities in the current discourse—matters. However, I propose that *le/Ø* and *se/ni* actually generate fundamentally different kinds of alternatives: one with alternatives of different properties being available (e.g., {man₁, woman₄, dog₇, ...}); and one with alternatives necessarily sharing the same property (e.g., {man₁, man₂, man₃, ...}).

Reexamining the definitions for *le/Ø* and *se/ni* in § 4, it is clear that neither requires that there actually exist alternatives—either in the domain of discourse, or in the world at large. For instance, if there only exists one individual α in the world w , the set of alternatives generated by $[[se \alpha]]^{w,g}$ will be singleton:

- (53) In world w , \exists one dodo.
 a. $[[se \text{ dodo}]]^{w,g} = \{x \mid x \text{ is a dodo in } w\}$
 $= \{ \text{dodo}_1 \}$

If the entity ‘dodo₁’ also exists in the discourse, then $[[le_{\delta,i} \text{ dodo}]]^{w,g}$ will generate the same singleton set. However, because of the presupposition on *le*, it makes for a stronger statement and will be preferred, all else being equal.

On the other hand, if world *w* contains billions of men, not all of whom are in the discourse, $[[le_{\delta,i} \text{ man}]]^{w,g}$ will always generate a subset of $[[se \text{ man}]]^{w,g}$; the former contains the contextually relevant men, and the latter contains all men in *w*, with no reference to the set of men salient in the discourse at all. Yet speaker commentary and speaker insistence on context indicate that this is not the case. Rather, *se*-man and other *se* α -type nominals in episodic sentences *imply* the existence of other such individuals with property α (or ‘man’) within the contextual domain $g(\delta)$. That is, in episodic sentences at least, *se/ni* alternatives are still ultimately constrained by context—just like *le/Ø* alternatives.

Thus, even if there clearly exist many sharks in the world, (49) (reproduced below in (54)) sounds unacceptable to speakers out-of-the-blue (i.e., with no salient set of discourse referents containing sharks). On the other hand, it becomes much more acceptable when a context involving multiple sharks in the discourse is invented. I paraphrase the context described by the speaker in (51)

- (54) a. ?? Sa fafaga e tama:loloa **se** malie.
 PST feed ERG man.pl **SE** shark
 Intended: ‘The men fed a shark.’
- b. Acceptable with the following context:
 The men are supposed to feed someone’s shark; they feed *some* shark, but evidently it was the wrong shark.

Relatedly, speakers often translate *se* α as ‘one of the α s’, implying the existence of other alternative entities with property α in the set of discourse-relevant entities. The speaker in (55) explicitly states this intuition:

- (55) Context: You are in the street, and a lady suddenly shouts that a man stole her bag.

Na vala’au le tama’ita’i ua gaoi e { **le** / **se** } tama:loa lana atou.
 PST shout LE lady PST.PERF steal ERG **LE** **SE** man 3sg.POSS bag

‘A lady shouted that a man stole her bag.’

Speaker: “[*Le/Ø* means] it was a *man* who stole the bag, or a particular man. But, *e se tama:loa* is more like ‘one of the men?’” . . .

MKS: “So, it should be that there are multiple men in the crowd?”

Speaker: “Yeah. We identify the crowd of men, and one of them.”

The comments in (55) also provide insight into the nature of the difference between *le/Ø* and *se/ni*: *le tama:loa* emphasizes the type of entity (*man*) or picks out a particular man whose existence is established in the discourse, and *se tama:loa* implies that alternative men exist. While *le tama:loa* and *se tama:loa* output the same truth value here, I propose that the use of article changes the sets of alternatives being compared. That is, *le tama:loa* produces a set of alternatives that can vary based on property (or “type”), while *se tama:loa* produces a set of alternatives that share the same property but vary based on identity (or “entity”).

Thus, the embedded clause in (55) might consist of the following alternatives, depending on the article used:

- (56) Context: “You are in the street. . .”
 Set of entities available in discourse = $g(\delta)$ = the set of all entities who might be on the street
- a. $[[ua \text{ gaoi e } \mathbf{le} \text{ tama:loa lana atou}]]^{w,g}$
 = { man₁ stole her bag, woman₄ stole her bag, dog₆ stole her bag, . . . }
- b. $[[ua \text{ gaoi e } \mathbf{se} \text{ tama:loa lana atou}]]^{w,g}$
 = { man₁ stole her bag, man₂ stole her bag, man₃ stole her bag, . . . }

Informally speaking, when speakers consider the alternatives present in the *le*-version (56a)—essentially, every entity present on the imagined street—they choose to emphasize the property ‘man’ in their interpretation, since this property might be unique among the alternatives. In the *se*-version (56b), however, all of the alternatives must be men, so ‘man’ can no longer be a noteworthy property of the referent. *Se tama:loa* simply refers to “one of the men”.

Evidence for a similar contrast exists in the related language Māori, and explored by Polinsky (1992), Bauer et al. (1997), and Chung and Ladusaw (2004), who compare the articles *tētahi* and *he*. Each analysis claims that *tētahi* somehow emphasizes the “individuation” of an entity, while *he*-marked nominals are type-based or “nonindividuated”. A personal favorite example illustrating this intuition comes from Chung and Ladusaw, who took advantage of a trip to the Te Papa Tongarewa Museum of New Zealand.

A more minimalist kind of discourse can be found in the museum, on the identifying cards or labels accompanying objects on exhibit. . . . When an object is exhibited along in a display case or is the only object of its type on display, then it is typically labeled by an indefinite headed by *he*, not *tētahi* (e.g., *He Mere* ‘Short Club Weapon’, *He Tiki* ‘Greenstone Pendant’). On the other hand, when several objects of the same type are exhibited together in a display case, the entire display, or each object in it, is typically labeled by an indefinite headed by *tētahi*, not *he* (e.g., *Ētahi Tiki* ‘Greenstone Pendants’, *Tētahi Tiki* ‘A Greenstone Pendant’ (in a display of several such pendants)’).

Chung and Ladusaw (2004:68-69)

Chung and Ladusaw conclude that “*tētahi* indefinites are composed as (relevant but arbitrarily selected) individuals”, while *he* indefinites are composed as properties” (69). While they explain these generalizations with their proposed compositional rules Restrict and Saturate outside of an alternative-semantics framework, the generalizations described here nevertheless draw clear parallels with *se/ni* and *le/∅*. Both *tētahi* and *se/ni* result in emphasis on entity, picking out one such α among a group of α s—in fact, *tētahi* is apparently morphologically composed of the definite article *te* and the numeral *tahi* ‘one’ (at least etymologically; see Chung and Ladusaw (2004:27-28)), which compares nicely with speaker interpretations of *se* α as “one of the α s”. Meanwhile, both *he* and *le/∅* result in emphasis on property or type.⁷

5.3 Formalizing the distinction

Up to this point I have only impressionistically sketched out this new component of the analysis of *le/∅* and *se/ni*. In this section, I begin to formalize these intuitions about the article system, stopping short of defining concrete denotations for each article. I explain why the present framework makes defining *le/∅* particularly challenging.

Despite the importance of context outlined above, it is clear that *se/ni* still has some sort of domain-widening effect. This will explain why negation and other polarity-related sentences with *se/ni* can be uttered without context, out-of-the-blue, with no discourse referents available. For instance, (57) does not require that there exist alternative queens of America in the discourse—in fact, it explicitly negates the possibility of any such entities.

- (57) E leai se tupu tama’ita’i o Amerika.
 NPST not.exist SE king lady GEN America
 ‘There is no queen of America.’

An equivalent episodic sentence—e.g., ‘SE queen of America was born’—would meanwhile require that alternative queens also exist in the world (and in the domain of discourse).

⁷Incidentally, the descriptive literature tends to label *tētahi* as “specific” and *he* as “nonspecific”—the inverse of what we would expect if the parallel to Samoan articles were perfect. However, as has already been made clear, descriptive names can be vague or even erroneous; for example, *tētahi* and *he* have also been described as representing contrastive definiteness (Biggs 1969). In any case, my goal is not to claim that the semantics of Māori and Samoan articles are identical, as the empirical landscape is far more nuanced. Rather, I only intend to show that the same sort of deep linguistic intuition is present in both languages, and could plausibly be present in a common ancestor.

In their analysis of German indefinites, Kratzer and Shimoyama (2002) posit various operators, structurally higher than the indefinites themselves, which render large numbers of alternatives interpretable by outputting singleton sets (or propositions). For instance, the operator \neg might sit wherever syntactic negation is located, and reduce a set of many propositions to one. If no such syntactically obvious operators exist, a covert “assertoric” operator can sit in a structurally high position—perhaps spec-CP—as a sort of “catch-all”, ensuring that sentences with no \neg , Q, \forall , etc. operators can be interpreted. One such type of sentence that requires an assertoric operator would be an episodic sentence.

The assertoric operator might look something like the following:

- (58) For $[[\alpha]]^{w,g} \subseteq D_{st}$:
- a. $[[\exists\alpha]]^{w,g} = \{ \lambda w'. \exists p [p \in [[\alpha]]^{w,g} \ \& \ p(w') = 1] \}$
 = the singleton set containing the proposition that is true in all worlds in which some proposition in α is true

(58) simply outputs a set containing a single, true proposition α . Let us also assume that the assertoric operator comes with a condition, described below:

- (59) *Non-trivial Alternative Generation*:
 There must exist discourse-available alternatives.
 $(| [[\alpha]]^{w,g} | > 1)$

If we say that the assertoric operator sits in spec-CP, above operators like \neg and in place of other spec-CP operators (like the Q operator), it is reasonable to assume that \neg and Q will compose with the set of alternatives before or instead of \exists . Thus, a *se/ni*-marked nominal in a negated sentence need not be restricted by Non-trivial Alternative Generation, and the nominal need not have existing alternatives in the discourse, nor does it itself need to exist.

However, in an episodic sentence containing *se/ni* (or *le/Ø*), there must exist alternatives in the domain of discourse. In the case of *se/ni*, there must be exist alternatives with the same property, which is why out-of-the-blue *se/ni*-nominals sound odd to speakers, and why adding a context in which it is likely that other such nominals exist makes it easier for speakers to accommodate the *se/ni* sentences.⁸

Why, however, do out-of-the-blue episodic sentences containing *le/Ø* sound acceptable? Or for that matter, why would any sentence with a one-of-a-kind *le*-marked nominal, with no existing alternatives with the same property, be acceptable? For instance, if the last dodo on earth has just been discovered, speakers can licitly talk about *le dodo*—even though there are no alternative dodos.

This is precisely where the “type” vs. “entity” distinction can assist. As argued in § 5.2, it is not the case that *le/Ø*-marked nominals generate alternatives with the same property; rather, it simply generates all alternative entities in the discourse. Thus, as long as it can reasonably be assumed that there exist discourse-salient entities—which should be the case in any world in which there is more than one entity—there will always be alternatives. Simply put, Non-trivial Alternative Generation is trivially satisfied by virtue of the fact that the speaker and addressee do not live in a void.

⁸One might wonder why non-polarity-related environments, such as clauses embedded under reportative verbs and nominals modified by relative clauses, allow for *se/ni*-marked nominals. I believe “out-of-the-blue” nature of these environments is greatly reduced as the sentence grows in complexity; that is, some imaginary context is more readily constructible by the addressee, so he is more likely to accommodate the statement. This is true even with unstressed *some* in English:

- (1) a. Some dog ate the food.
 b. I am certain that some dog ate the food.
- (2) a. Some dog ate the food.
 b. Some lost dog ate the food.

To my ear, and according to a rather superficial survey I conducted with peers, the more complex sentences in the (b) examples sound more acceptable out-of-the-blue than the (a) examples. A speaker can, after all, only use relative clauses and adjectives on a nominal felicitously if it is presupposed that there exist other such nominals which do not possess those properties. In the case of embedded clauses, speakers may simply be more willing to imagine the more detailed sentence as part of a larger narrative.

Unfortunately, an actual denotation for le/\emptyset under the Hamblin framework is difficult to pin down. On one hand, the denotation must generate a set of all discourse entities of any property; on the other hand, the property with which le/\emptyset composes must somehow be preserved. Thus, though *le tama:loa* ‘LE man’ might be represented by $\{\text{man}_1, \text{woman}_2, \text{dog}_7\}$, the fact that the nominal being described has the property ‘man’ must ultimately be recoverable—otherwise, *Sa siva le tama:loa* ‘the man danced’ might mean that the woman danced, the dog danced, or that any other entity in the discourse danced.

I leave the precise denotation of le/\emptyset for future work. Even so, the intuition behind my analysis of the article dichotomy is complete: se/ni -marked nominals generate alternatives with the same property from an extremely wide domain, while le/\emptyset -marked nominals generate alternatives of any property from within the discourse. An assertoric operator then outputs a single proposition, as long as alternatives to that proposition existed. For le/\emptyset , this will always be the case. For se/ni , however, this will only be the case if there exist entities with the same property in the discourse. If, for instance, se/ni is instead composed by a different operator (questions, negation, etc.), it has a domain-widening effect not restricted by the discourse.

6 Conclusions

This article has examined the complexities of the Samoan article system, starting with a comparison of the system with the semantics of German indefinites, and concluding that such a comparison—even supplemented by a presupposition on le/\emptyset —is insufficient to capture some of the nuances behind the distribution of se/ni .

The intuitions behind the interpretational and distributional differences of le/\emptyset and se/ni are clear: le/\emptyset generates alternative entities in the discourse of any kind, while se/ni generates alternative entities of the same kind from anywhere in the world. This explains why se/ni appears to have a domain-widening effect; this in turn explains why le/\emptyset is often given wide-scope or referential readings by speakers, and why speakers interpret se/ni -marked nominals as conveying a level of uncertainty as to which entity of the property in question participated in an event. le/\emptyset additionally possesses a presupposition, ensuring that it can only be interpreted if the nominal in question actually exists in the world. This explains some of the restrictions on its distribution (e.g., never after *leai* ‘does not exist’), as well as another component of uncertainty that surrounds se/ni : uncertainty regarding the existence of the nominal and the occurrence of the event in question.

Finally, there exists some operator catch-all operator in a structurally high position which, if all else fails, composes the sets of alternatives generated throughout the course of a derivation. However, this operator bears a condition which requires alternatives to have to been generated, and for said alternatives to exist in the domain of discourse. This is almost trivially satisfied by le/\emptyset -containing sentences, since le/\emptyset -marked nominals can be of any property, and speakers will likely assume or accommodate the existence of other entities in the discourse. However, this condition is not so easily satisfied by se/ni -containing sentences. Because se/ni only generates alternatives of a certain property, it may be the case that the discourse has not supplied evidence for the existence of such other entities. Thus, speakers cannot as easily accommodate se/ni in out-of-the-blue (i.e., “context-less”) sentences that are episodic (i.e., that have no other operator to compose them). Meanwhile, le/\emptyset -marked nominals can exist in most any environment, save the few that conflict with its presupposition.

The work that remains to be done is a proper formalization of the intuitions expressed here. While the denotation of se/ni is relatively simple, requiring no adjustments from the Hamblin-style denotation of *irgendein* from Kratzer and Shimoyama (2002), le/\emptyset is more challenging. le/\emptyset -marked nominals must simultaneously denote a set of entities with different properties, while compositionally preserving the single property asserted by the speaker. I leave this particular issue for future efforts, but suggest that a “two-dimensional” alternatives framework, of the sort in Rooth (1992), might provide a more straightforward path to a denotation.

Other work which remains to be done, which I have not elaborated on here, is the semantic effect of plurality on the two articles under discussion. While I have thus far discussed le/\emptyset and se/ni as two completely distinct and internally consistent items, it is likely that \emptyset and *ni*, the plural versions of

each article, have special semantics of their own not accounted for here. Because my elicitations have not led me to believe there are significant interpretational differences between the plural and singular forms of these articles, and because past data (such as the grammar) do not hint at such differences, I have sought to first thoroughly explain one of the two dichotomies in the article system: the property of “specificity”. I have also incorporated data from both plural and singular articles into my claims, although I have primarily relied on singular data. A complete semantics of Samoan *le/Ø* and *se/ni*, however, would systematically uncover the effects of plurality on interpretation, and separate these from any potential effects on “specificity”.

Despite these shortcomings, the data here are theoretically promising, and I believe merit continued investigation—perhaps with more speakers, and perhaps on the island of Samoa itself. I conclude with hopes to proceed in this analysis and incorporate Samoan into the broader discourse on the semantics of indefinites, alternatives, and specificity.

7 Works cited

- Bauer, W., Parker, W., Evans, T. K., & Teepa, T. A. N. (1997). *The Reed Reference Grammar of Māori*. Reed.
- Biggs, B. (1969). *Let's Learn Maori: A Guide to the Study of the Maori Language*. Reed.
- Carlson, G. N. (2005). Generics, habituals, and iteratives. In *The Encyclopedia of Language and Linguistics* (2nd ed.). Elsevier.
- Chierchia, G. (2013). *Logic in Grammar*. Oxford University Press.
- Chung, S., & Ladusaw, W. A. (2004). *Restriction and Saturation*. The MIT Press.
- Clemens, L., & Massam, D. (2021). Polynesian languages and their contributions to theoretical linguistics. In L. Clemens & D. Massam (Eds.), *Polynesian Syntax and its Interfaces* (pp. 1–10). Oxford University Press.
- Collins, J. N. (2014). The distribution of unmarked cases in Samoan. In I. W. Arka & N. L. K. M. Indrawati (Eds.), *Argument Realizations and Related Constructions in Austronesian Languages*. Asia-Pacific Linguistics.
- Cook, K. W. (1996). The Cia suffix as a passive marker in Samoan. *Oceanic Linguistics*, 35(1), 57–76.
- Diesing, M. (1992). *Indefinites*. MIT Press.
- Enç, M. (1991). The semantics of specificity. *Linguistic Inquiry*, 22, 1–25.
- Farkas, D. (1994). Specificity and scope. In L. Nash & G. Tsoulas (Eds.), *Langues et grammaires* (pp. 119–137).
- Falauş, A. (2013). Introduction: Alternatives in semantics and pragmatics. In *Alternatives in Semantics* (pp. 1–35). Palgrave Macmillan.
- Giannakidou, A. (2011). Negative and positive polarity items. In K. von Stechow, C. Maienborn, & P. Portner (Eds.), *Semantics: An International Handbook of Natural Language Meaning* (pp. 1660–1712). De Gruyter Mouton.
- Groenendijk, J. A. G., & Stokhof, M. J. B. (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers* (Doctoral dissertation). Universiteit van Amsterdam.
- Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language*, 10(1), 41–53.
- Heim, I. (1991). Artikel und definitheit. In A. V. Stechow & D. Wunderlich (Eds.), *Semantics: An International Handbook of Contemporary Research* (pp. 487–535). Mouton de Gruyter.
- Hohaus, V. (2018). Quantification in Samoan at the syntax-semantics interface. *Presented at the Workshop on the Syntax of Polynesian Languages at the University of Toronto*.
- Ionin, T. (2006). This is definitely specific: Specificity and definiteness in article systems. *Natural Language Semantics*, 14, 175–234.
- Koopman, H. (2012). Samoan ergativity as double passivization. In L. Brugé, A. Cardinaletti, G. Giusti, N. Munaro, & C. Poletto (Eds.), *Function Heads: The Cartography of Syntactic Structures*. Oxford University Press.
- Kratzer, A. (2005). Indefinites and the operators they depend on: From Japanese to Salish. In G. N. Carlson & F. J. Pelletier (Eds.), *Reference and Quantification: The Partee Effect* (pp. 113–142). CSLI Publications.

- Kratzer, A., & Shimoyama, J. (2002). Interderminate pronouns: The view from Japanese. In Y. Otsu (Ed.), *Proceedings of the Third Tokyo Conference on Psycholinguistics* (pp. 1–25). Hituzi Shobo.
- Krifka, M. (1995). The semantics and pragmatics of polarity items. *Linguistic Analysis*, 25, 209–257.
- Montague, R. (1970). English as a formal language. In B. Visentini (Ed.), *Linguaggi nella società e nella tecnica* (pp. 189–223). Edizioni di Comunità.
- Mosel, U., & Hovdhaugen, E. (1992). *Samoan Reference Grammar*. The Institute for Comparative Research in Human Culture; Scandinavian University Press.
- Polinsky, M. (1992). Maori *he* revisited. *Oceanic Linguistics*, 31, 229–250.
- Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, 1(1), 75–117.
- Tollan, R. (2018). Unergatives are different: Two types of transitivity in samoan. *Glossa: A Journal of General Linguistics*, 3(1), 1–41.