Fake reefs are sometimes *reefs* and sometimes not, but are always compositional

Hayley Ross, Najoung Kim & Kathryn Davidson[*]

Abstract. In semantics, adjective modification is typically handled with set intersection, such that $[\![\text{yellow flower}]\!] = [\![\text{yellow}]\!] \cap [\![\text{flower}]\!]$. Thus a *yellow flower* is a *flower*. Such an account, however, runs into problems for adjectives like *fake* or *counterfeit*, which display a privative entailment: a *fake fire* is not a *fire* and a *counterfeit dollar* is not a *dollar*. Moreover, recent work shows privativity cannot easily be encoded as a property of specific adjectives like *counterfeit*, since e.g. *counterfeit watch* robustly licenses the subsective entailment of being a *watch* (Martin 2022). We gather judgments on over 300 adjective-noun bigrams (of which 57 novel, i.e. zero corpus frequency), and show that privativity depends on the adjective, noun and context, and can be manipulated for the very same adjective-noun bigram by presenting it in different contexts. This poses a challenge for theories which fix privativity as a property of the adjective and always use the same method of composition (Partee 2010, del Pinal 2015). Moreover, we find no difference in participant behavior between novel adjective-noun bigrams and high frequency ones, suggesting that the process is nonetheless compositional and not the result of convention or memorized idiosyncrasy. Our results support compositional accounts like Martin (2022) (which modifies del Pinal 2015) and Guerrini (2022a), which treat privativity as context-dependent.

Keywords. adjectives; nouns; compositionality; privativity; entailment; semantics; experimental semantics

1. Introduction. A central concern for the study of meaning is how the meanings of complex expressions are composed from the meanings of their constituent parts. The fact that people understand completely novel phrases provides an argument that meaning must be governed by some version of compositionality – that the meaning of phrases is a function of their parts (perhaps including context as one part) and how they are syntactically combined (Partee 2009). This paper, following in a growing tradition (Partee 2009, 2010, Szabó 2012, del Pinal 2015, i.a.) studies the dynamic interaction of meaning and context through the lens of (privative) adjective modification. So-called "privative" adjectives like *fake* cause particular issues for a compositional account since they seem to license the inference that e.g. a *fake fire* is not a *fire* at all. Let us begin with the way that most semantics textbooks treat the modification of nouns by adjectives, namely as simple set intersection (Heim & Kratzer 1998, Coppock & Champollion 2023).[1]

(1) $[\![\text{yellow flower}]\!] = [\![\text{yellow}]\!] \cap [\![\text{flower}]\!] = \{x : x \text{ is yellow}\} \cap \{x : x \text{ is a flower}\}$

[1]To their credit, both textbooks proceed to mention how gradable and/or nonintersective adjectives are not adequately addressed by their account.

This predicts an *intersective* entailment, where the resulting composed concept is both an instance of the adjective and an instance of the noun (a *yellow flower* is both *yellow* and *a flower*). Set intersection-based accounts, however, quickly run into problems. While they works for simple cases like *yellow flower*, many adjectives yield *subsective* or even *privative* patterns, where the resulting concept isn't an instance of the adjective or the original noun at all.

(2) *Intersective inference*
This is a yellow flower.
∴ This is yellow.
∴ This is a flower.

(3) *Subsective inference*
This is a small skyscraper.
∴ This is a skyscraper.
∴̷ This is small.

(4) *Privative inference*
This is a fake fire.
∴ This is not a fire.

In other words, if we want to hold that *fake fires* are not in the set of *fires*, then no amount of intersection or subsection with that set of *fires* will ever yield *fake fires* – something more sophisticated is required. Further, Martin (2022) shows with experiments that which entailment is available is actually specific to the adjective-noun combination, not a property of the adjective, contra previous literature (Kamp & Partee 1995, Partee 2010 i.a.). For example, *counterfeit* may license a privative or subsective (likely even intersective) entailment depending on the noun (and accompanying context):

(5) *Subsective/intersective inference*
This item is a counterfeit watch.
∴ This item is a watch.
∴? This item is counterfeit.

(6) *Privative inference*
This is a counterfeit dollar.
∴ This is not a dollar.

This per-adjective and per-noun variation raises the question of whether these adjective-noun combinations and their entailments are computed fully compositionally on the fly (based on just the given adjective, noun and context), or whether there is an element of convention or past experience necessary to derive these varying entailments, perhaps simply storing (memorizing) the entailments for each bigram. A significant body of processing work (Arnon & Snider 2010, Tremblay & Baayen 2010, Caldwell-Harris et al. 2012, O'Donnell 2015 i.a.) reveals plenty of cases where humans *don't* appear to compose meaning (or even morphemes) on the fly: chunks of various sizes from multi-morpheme words to entire idiomatic expressions, especially highly frequent words or expressions, can get stored as units and trigger priming effects in experimental studies, whether their meaning is idiomatic or fully compositional from their parts. Given this ability to store complex expressions, and given the difficulty in treating these adjectives with straightforward set intersection/composition, one might propose that words with widely varying, use-case-specific meanings (and resulting entailments) like *fake* are primarily handled by storing (memorizing) their most frequent use cases, i.e. storing entire adjective-noun bigrams. In that case, true composition of these adjectives into novel adjective-noun pairs, such as *fake scarf* or *fake reef* might be difficult or result in widely varying results between people (especially when varying contexts are taken into account). This paper explores the effect of experience (as measured by corpus frequency of the target bigram) and context on adjective-noun composition, especially for novel adjective-noun bigrams (essentially a "wug" test for adjective-noun composition).[2]

---

[2]This paper treats compositionality and memorization as two ends of the scale, but naturally there are possibilities

In this paper, we gather a large quantity of adjective-noun entailment judgements for both high-frequency and novel (zero corpus frequency) adjective-noun bigrams and show that entailments depend not just on the adjective but also on the noun and the context. Further, we show that novel adjective-noun bigrams and their entailments are handled as productively and consistently by participants as high-frequency ones, suggesting that despite the significant variation by adjective, noun, and context, a compositional account is still on the right track.

2. Background. Historically, privativity has been defined as an adjective-specific phenomenon which negates the noun that the adjective combines with. A *fake gun* is said to be precisely *not a gun*, i.e. not in the set of *guns*. This view means that we cannot derive the meaning of *fake gun* from the set of (real) *guns* as it stands through any set-theoretic operation (intersection, subsection or otherwise), and poses a significant problem for compositionality, which requires the meaning of a complex expression to be derived solely from the meaning of its constituent parts (Szabó 2012). Canonical examples of such "privative" adjectives include *fake, false, counterfeit, knock-off, mock, former* and perhaps also *artificial, pretend,* and *virtual* (see Nayak et al. 2014 for an attempt at comprehensive categorization). Further, Partee (2009) points out that a simple set complement approach to privativity, however implemented, cannot be on the right track. If (7a) is true because *fake gun* is in the complement of *gun*, then (7b) should not be well-formed. As Martin (2022) shows, participants are happy to use the plain noun *gun* or *bear* (7c) to refer to objects know not to be real; clearly the use of *gun* or *bear* is not as simple as a set complement approach assumes.

(7) a. A fake gun is not a gun.

   b. Is that gun real or fake?

   c. This [*pointing at a toy bear*] is a bear.

Further, Martin (2022) shows experimental evidence that inference patterns for so-called privative adjectives vary widely depending on the noun used, with some items, such as *fake exam*, being rated very high for counting as *exams*, while other items like *fake dollar* are rated very low (are not *dollars*). In other words, this is not an adjective-specific phenomenon, but rather depends at minimum on the noun (as well as on the context, as we will see later). Martin uses a simple and direct approach, which we will also adopt, asking people (with no further context): "Is a fake exam an exam?" This differs from previous work (Pavlick & Callison-Burch 2016a,b, Pustejovsky 2013) which asks participants to rate these inferences given a particular sentence context. For example, Pavlick & Callison-Burch (2016b) ask whether (8a) entails (8b). While more realistic than out-of-the-blue judgements, this also creates a much noisier picture, as demonstrated by this example: participants rate (8a) as contradicting (8b), but *denied* and the world knowledge of pharmacists selling medicine also seem to be driving part of this inference. It actually remains unclear whether the counterfeit medicine that the pharmacists were selling qualifies as medicine in this scenario.

---

that are intermediate between the two. In particular, it is easy to imagine an analogical approach where people memorize the inferences for a certain number of high-frequency bigrams, and then handle novel bigrams by analogy to them where possible. In this model, a participant might encounter *counterfeit scarf* as novel, but reason that it is subsective (is a *scarf*) by analogy to other clothing items and accessories such as *watch* or *handbag* or *sneakers* which they have seen *counterfeit* occur with subsectively. We hope to explore the validity of such a model in future work, though our results in this paper do not provide obvious support for such an account.

(8) a. Pharmacists in Algodones denied selling counterfeit medicine in their stores.

b. Pharmacists in Algodones denied selling medicine in their stores.

While Martin's experiment avoids these issues, Martin's choice to hand-pick 8 nouns for each of his 11 adjectives (of which 5 "privative": *artificial, counterfeit, fake, false* and *mock*) leaves many adjective-noun combinations unexplored and risks allegations of cherry-picking (as Martin himself notes). Moreover, context seems like it should play a role in determining when a *(fake) gun* counts as a *gun*, or a *(fake) watch* counts as a *watch*: a toy watch which does not tell the time may not qualify as a watch, but a fake Rolex watch might. While Martin is clearly aware of context in the earlier chapters of his dissertation, where he discusses the ambiguity between intersective and subsective readings of the same adjective-noun pair (e.g. *beautiful dancer*), it is not factored into his experiment on privativity. This paper addresses context directly by writing contexts which bias the target bigram towards either a privative or subsective reading.

3. Choice of adjective-noun bigrams. We set the stage for the experiments in this paper by establishing a cross of 38 nouns by 12 adjectives to test, such that a large proportion of the resulting bigrams are zero frequency in a large corpus, thus presumed novel for experiment participants.

3.1. SELECTION BY CORPUS FREQUENCY. We begin with 300 nouns which commonly occur with a wide range of adjectives (Pavlick & Callison-Burch 2016a), plus the 36 nouns used in Martin (2022), for a total of 328 unique nouns. We consider 6 "privative" adjectives of interest: *fake, counterfeit, false, artificial, knockoff* and *former*.[3] (Since we established in the introduction that such adjectives need not always be privative, from here on out the phrase "privative adjective" will refer to adjectives such as *fake* that have a tendency to result in privative entailments.) For our corpus, we use C4 (Colossal Clean Crawled Corpus), a corpus developed for natural language processing (Raffel et al. 2020, Dodge et al. 2021) which contains over 130 trillion words of English crawled from the web. For each of these privative adjectives, we count their frequencies in C4 and (hand) select 6 intersective/subsective adjectives as "controls" which each have a similar frequency to one of the privative adjectives and which have relatively few selectional restrictions: *useful, tiny, illegal, homemade, unimportant* and *multicolored*. Frequencies are shown in Table 1. In particular, *multicolored* is chosen since it is a low-frequency example of a colour adjective, classic examples of (supposedly) intersective adjectives, and *homemade* is chosen since it targets the manner of manufacture, similar to *counterfeit* and *artificial*, without being obviously privative.

Crossing these 12 adjectives and 328 nouns yields 3936 bigrams. Each of these bigrams' frequencies is then counted in C4. Using these counts, we algorithmically select 43 nouns from the original 328 nouns to yield a high proportion of zero-frequency bigrams when crossed with the target 12 adjectives (40% of bigrams), while excluding vague or ambiguous nouns like *part, stuff* (vague when out of context) and *bar* (ambiguous between two very different meanings).[4] This yields the nouns in (9), for a total of 516 bigrams.

---

[3]*Knockoff* in particular is included as one pilot participant reported that they would never use *counterfeit*, which is common in the literature on privative adjectives, and would use *knockoff* (*knock-off*) to refer to that concept.

[4]5 such nouns later excluded for being vague/ambiguous were originally included in this cross and were included in Experiment 1, for a total of 48 nouns and 576 bigrams.

Adjective	Tokens	Adjective	Tokens
former	15.8M	useful	13.6M
false	4.6M	tiny	5.8M
artificial	3.9M	illegal	4.5M
fake	3.1M	homemade	2.2M
counterfeit	450K	unimportant	170K
knockoff	57K	multicolored	93K

Table 1: Adjective frequencies in the C4 Corpus (130T words)

(9) accusation, act, air, car, cat, concert, couple, crowd, drug, fact, fire, form, gentleman, glance, gold, hand, head, home, house, idea, image, information, jacket, laugh, life, light, man, market, material, person, plan, pole, post, reef, report, scarf, sea, sign, sofa, war, water, work, world

The upside of a rigorous adjective-noun cross is that we can avoid the concerns of cherry-picking or missing out on possible unexpected patterns, as acknowledged in Martin (2022). The downside of a rigorous cross is that, especially for the typically privative adjectives, many of these combinations will not make sense: many of these adjectives have comparatively narrow selectional restrictions such as requiring a process of manufacture or a surface which can have a color. Combinations like *counterfeit accusation* or *homemade life* must be excluded before we can reasonably ask questions like "Is a counterfeit N still an N?" Experiment 1 performs this task by excluding all bigrams our participants find hard to assign a meaning.

3.2. EXPERIMENT 1. Experiment 1 is a simple design to filter out clearly nonsensical combinations resulting from the blind crossing of adjectives and nouns. Participants are presented with a bigram and asked "How easy is it to imagine what this would mean?", as shown in Figure 1.[5] A follow-up question designed to rule out ambiguous nouns was also included, but turned out not to be effective, so ambiguous nouns were ruled out manually instead. 144 native American English speakers[6] were recruited on Prolific (of which 7 were excluded due to failed attention checks and/or not meeting the criteria for native English speaker); the study was implemented in Qualtrics. Participants were paid pro rata at $12/hour; the experiment took 4 minutes on average. Each participant saw 14 questions (12 target bigrams + 2 attention check bigrams), such that each bigram received 3 ratings in total.

We categorize bigrams whose ratings were majority "very hard" or "somewhat hard" as nonsensical, and exclude them from subsequent experiments. This leaves 305 bigrams, of which 19% (57) are zero frequency in C4, i.e. almost certainly novel to new participants, and another 7% (20)

---

[5]Previous work studying novel adjective-noun combinations (Vecchi et al. 2017) uses a more complex pairwise ranking approach to precisely measure semantic deviance, but we are only concerned with filtering out bigrams which are obviously nonsensical, allowing a simpler question formulation.

[6]For the purposes of this paper, we define "native American English speaker" as having learned English before the age of 5, speaking English as their main language, and self-identifying as a speaker of American English. The concept of native speaker is not without issues and this is merely intended as a convenient way to expect shared language experiences among our participant sample.
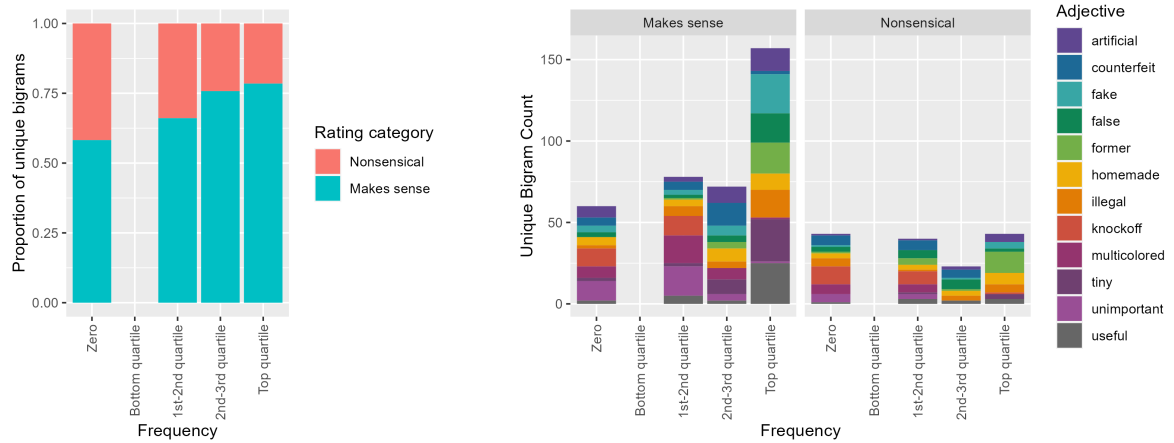
counterfeit scarf

How easy is it to imagine what this would mean?

| Very hard | Somewhat hard | Somewhat easy | Very easy |

Is it easy to think of multiple very different meanings?

| No | Yes |

Figure 1: Screenshot of an example question in Experiment 1

(a) Proportion of unique bigrams by frequency

(b) Unique bigrams by frequency and adjective

Figure 2: Bigram distribution after filtering in Experiment 1 . Bigrams with majority rating "very hard" or "somewhat hard" are categorized as nonsensical, the rest as making sense.

are very low frequency (bottom quartile), so also quite possibly novel to participants. The relationship between frequency and rating categorization is shown in Figure 2. Interestingly, some top quartile[7] bigrams are rated as nonsensical in this setting. This is mostly explained by a side effect of how we count bigrams: in the interest of computational efficiency, we count an occurrence of e.g. *knockoff hand*, *tiny water*, *former concert* or *homemade air* (all top quartile but rated nonsensical) whenever that string occurs in the text, even if it actually occurs as part of *knockoff handbag*, *tiny water droplet*, *former concert manager* or *homemade air freshener*. In isolation, outside of these noun-noun compounds or other longer expressions that contain them, these bigrams do not make sense.

---

[7]Here and in the rest of the paper, quartiles and percentiles are calculated over the 3936 total bigrams from Section 3.1, not over every possible corpus bigram, which would be computationally intractable. Thus, terms like "high frequency" or "top quartile" should be interpreted in relative rather than absolute terms.

Is a counterfeit scarf still a scarf?

Definitely not   Probably not   Unsure   Probably yes   Definitely yes

Is a fake gun still a gun?

Definitely not   Probably not   Unsure   Probably yes   Definitely yes

Figure 3: Screenshot of two example questions (shown on separate pages) in Experiment 2

4. Experiment 2: *Is an A N an N?*.

4.1. METHOD. Experiment 2 uses the same design as Martin (2022), with one slight modification, and asks participants *Is an A N still an N?* for each of the 305 adjective-noun bigrams left after filtering in Experiment 1. The goal of this experiment is twofold: to show (for a wider range of adjectives and nouns than Martin 2022) that privativity varies depending on the noun, and to investigate whether participants behave differently for high and zero frequency (assumed to be novel) bigrams. The only modification to Martin's experiment is that the adverb *still* is included in the question phrasing: the experiment designers felt that the questions were more natural with *still* included.[8] 510 native American English speakers were recruited on Prolific (of which 15 were excluded due to failed attention checks and/or not meeting the criteria for native English speaker); the study was implemented in PCIbex (Zehr & Schwarz 2018). Participants were paid pro rata at $12/hour; the experiment took 3 minutes on average. Each participant saw 12 questions of the form *Is an A N still an N?* (4 typically-intersective adjectives, 4 typically-privative adjectives, 4 fillers), for a total of 10+ ratings/bigram.[9] An example question is shown in Figure 3. Lastly, since some bigrams which may not make sense to everyone likely remain after Experiment 1, we explicitly alert participants in Experiment 2 to this possibility and instruct them to use the "Unsure" rating if a combination does not make sense to them.

4.2. RESULTS. Mean bigram ratings are shown in Figure 4, and the full set of data is shown in Figure 5. We find that each so-called "privative" adjective in fact yields graded variation from privative to subsective depending on the noun, with ratings spanning all the way from 1 ("Definitely not [an N]") to 5 ("Definitely yes [an N]"). In Figure 5, we also notice a high variance for items with means in the centre of the scale, suggesting that it's often not the case that most participants

---

[8]In this case, *still* seems to be targeting the intuition that for most of these cases, a combination such as *fake scarf* or *unimportant sign* might *not* be a scarf or sign, or conversely might be a scarf *despite* being fake. Arguably, this is a case of *Maximize Presupposition* (Heim 1991). Including *still* should not cause problems for the experiment by presupposing the question being researched, since it merely presupposes that it is possible for some adjective-noun combinations to be privative, a fact which participants are most likely aware of. There is a possibility that it may yield a temporal interpretation when combined with *former* (*Is a former house still a house?*), but this is still compatible with the question interpretation we intend.

[9]Due to some issues with PCIbex, the experiment did not yield an even number of ratings per bigram. In the analysis of this experiment, we randomly sample and cap the number of ratings at (10-)12 ratings/item.
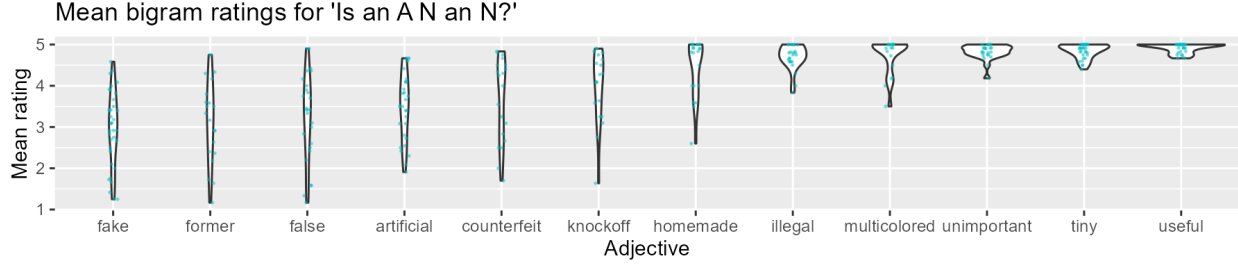
Figure 4: Mean bigram ratings for Experiment 2, where 1 is most privative and 5 is most subsective.

judge these as "somewhat an N" but rather that different participants rate them as mostly subsective or mostly privative respectively. We also find that "subsective" adjectives are usually subsective (ratings at ceiling for being an instance of N), warranting the name, but are nonetheless not so clearly subsective with certain nouns (e.g. *homemade cat, multicolored gold, illegal gentleman*).

Secondly, we investigate whether frequency has an effect on rating variance. For example, if we think that these entailments are adjective-noun specific to the point that they are conventionalised and (at least partially) memorized, participants might have a stable, conventionalised meaning or entailment only for high-frequency bigrams, and be unsure about or disagree on the meaning or entailment for novel bigrams, leading to high variance. In fact, we find no effect of frequency on rating variance, as shown in Figure 6. A linear regression in R (R Core Team 2023) shows that frequency correlates poorly with the variance in the ratings (typically subsective: $R^2 = 0.009$ / adjusted $R^2 = 0.002$, typically privative: $R^2 = 0.014$ / adjusted $R^2 = 0.008$, both $p > 0.05$). Instead, participants agree to a similar degree on the meaning and inferences for high-frequency and novel adjective-noun bigrams. Some very high frequency bigrams such as *artificial tree, former house* or *fake plan* show high variance in ratings ($\mu = 3.50, \sigma^2 = 1.83$; $\mu = 3.63, \sigma^2 = 1.76$; and $\mu = 3.50, \sigma^2 = 2.50$ respectively), suggesting that these bigrams do not have a conventionalized meaning/entailment when presented out of the blue. Moreover, some zero frequency bigrams like *knockoff image* and *counterfeit scarf* have quite low variance ($\mu = 4.90, \sigma^2 = 0.10$ and $\mu = 4.80, \sigma^2 = 0.18$), showing that participants compose even novel bigrams systematically.[10]

4.3. DISCUSSION. The results from Experiment 2 clearly show that no adjective always results in privativity, lending further weight to the findings in Martin (2022) and dispelling any concerns about cherry-picking. Each "privative" adjective may be rated as either subsective, privative, or somewhere in between depending on the noun it is presented with. We further see relatively high variance between participants for some high frequency bigrams and low variance for some novel bigrams, suggesting that high frequency does not correlate with a fixed conception of bigram meaning or entailment, and that previous exposure to a (potentially) privative adjective-noun pair is not required to draw this inference, even for adjectives with relatively broad meanings like *fake*. In-

---

[10]The adjectives *knockoff* and *counterfeit* have the lowest-variance novel bigrams; *fake*, which generally modifies a broader range of aspects of a noun, starts at $\mu = 2.00, \sigma^2 = 0.89$ with *fake reef*, which is still relatively low variance for this experiment. Variances for individual items reach as high as $\mu = 2.91, \sigma^2 = 3.29$ for *former image*.
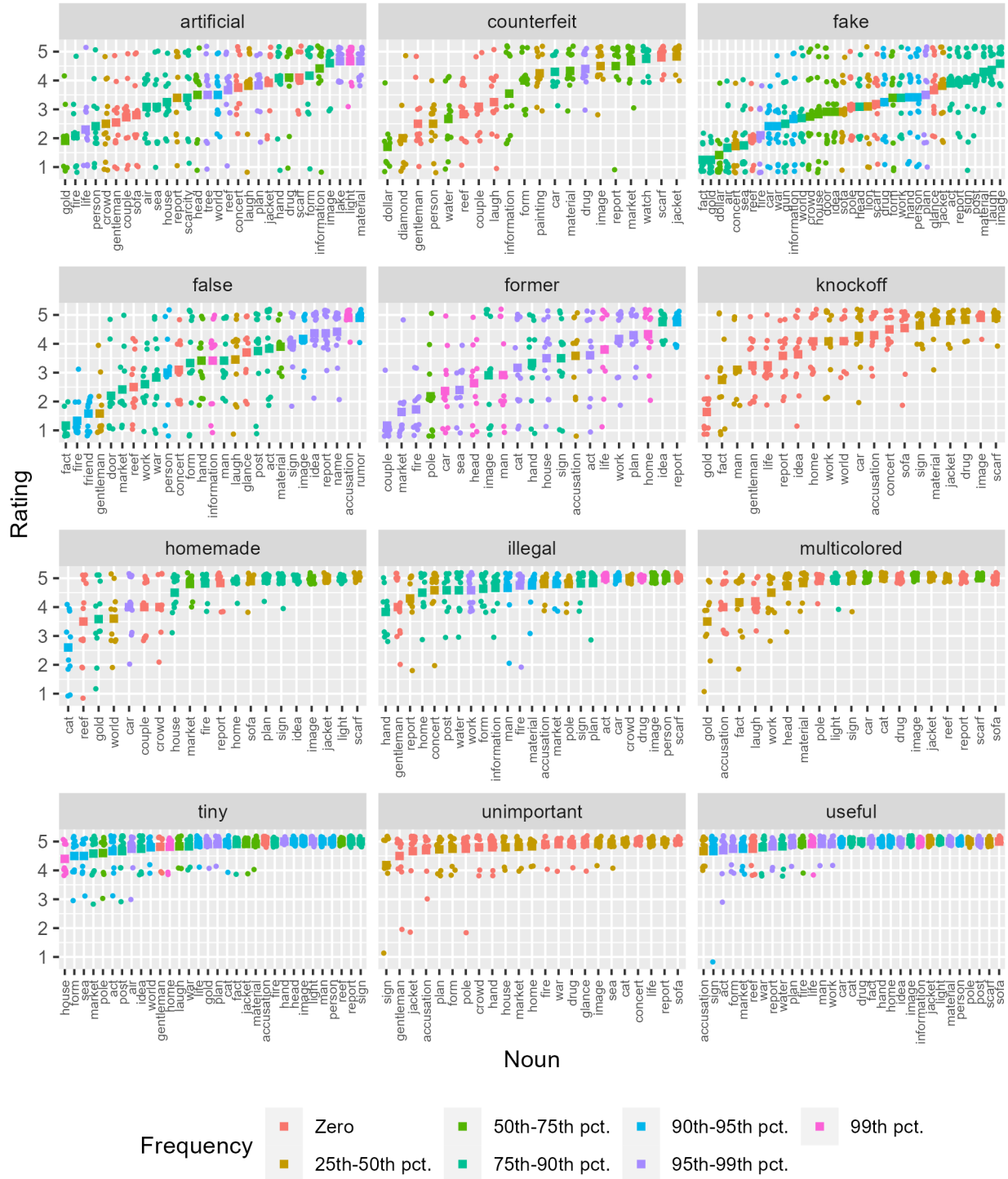
Figure 5: Ratings for all participants for all 305 bigrams in Experiment 2, where 1 is most privative and 5 is most subsective. ■ shows the mean for each bigram.
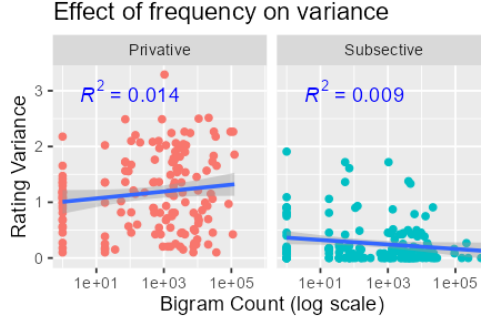
Figure 6: No effect of frequency in Experiment 2.

A political party disguises a fundraiser as a concert so that they can hold it at a venue where political rallies aren't allowed. They even hire an up-and-coming band to sing at the event. The false concert is a great success and the attendees enjoy the music as well as networking with the political candidates.

In this setting, is the false concert still a concert?

Definitely not	Probably not	Unsure	Probably yes	Definitely yes
○	○	○	○	○

A well-known band gets into trouble when it emerges that they included a false concert in their tax returns, which they claim had huge financial losses (letting them get away with paying very low taxes), but which never actually happened.

In this setting, is the false concert still a concert?

Definitely not	Probably not	Unsure	Probably yes	Definitely yes
○	○	○	○	○

(a) Subsective-biased context   (b) Privative-biased context

Figure 7: Screenshots of contexts for *false concert* in Experiment 3.

stead, we suspect that a large proportion of the high variance is caused by participants imagining different contexts for the bigrams (which were presented out of the blue in Experiment 2), such that e.g. *fake* might target different aspects of the noun's properties or different properties might be relevant for noun-hood in that context. For example, a *fake door* might qualify as a *door* on a theatre stage, but not in a house if it's a painted picture of a door that doesn't actually open.

We also find tentative evidence suggesting that there is no clear split between (sometimes) privative adjectives vs. always subsective adjectives, since some supposedly non-privative adjectives like *homemade* or *multicolored* do not receive ratings at ceiling for all nouns. A *homemade cat* is considered not or probably not a cat by over half the participants, and this is likely to become even more apparent if we supply a context where the cat must be able to breathe, purr and walk around. This "emergent" or "contingent" (noun-specific) privativity has previously been observed for constitutive material adjectives like *stone*, with the classic example of *stone lion* (Partee 2009, del Pinal 2015, Martin 2019); our results tentatively suggest that this phenomenon is not just limited to material adjectives. Future experiments are planned to illuminate the privativity of classically subsective adjectives, including material adjectives, in more detail.

5. Experiment 3: Context.

5.1. METHOD. For 12 adjective-noun bigrams from Experiment 2, we construct two contexts designed to bias the reader towards a subsective or privative entailment respectively. Two example contexts for *false concert* are shown in Figure 7. The contexts are manually written by the exper-
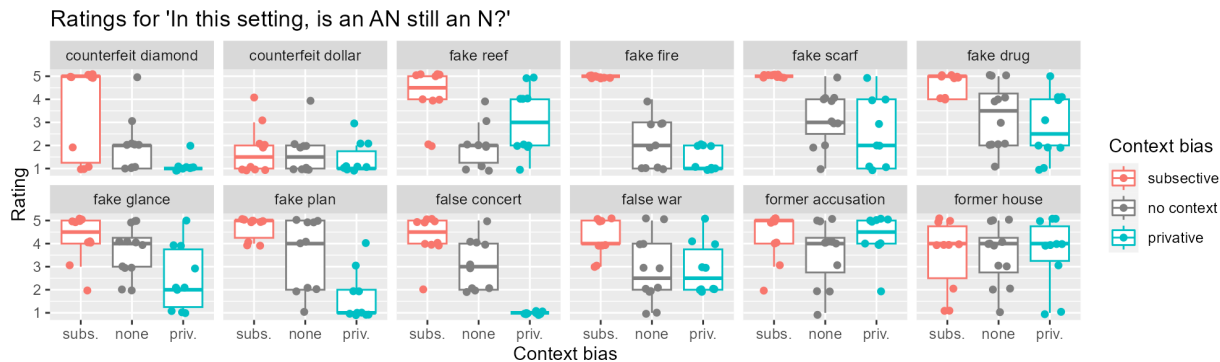
Ratings for 'In this setting, is an AN still an N?'

Figure 8: Results for Experiment 3, where 1 is most privative and 5 is most subsective. The ratings from Experiment 2 are shown in gray.

iment designers to yield an subsective or privative entailment according to their intuitions, with the intent that these entailments will be reproduced by the participants. (Note that this means that participants are presented with an instance of e.g. *fake reef* in a particular context and only get to judge whether this object described as a *fake reef* is a *reef* in this setting; they do not get to judge whether it is appropriate to call this object a *fake reef* in the first place.) Specifically, we select 6 pairs of adjective-noun bigrams from Experiment 2 with similar middling ratings and high variance, such that one bigram is zero/low frequency and the other is high frequency. This yields the pairs *counterfeit diamond/dollar, fake reef/fire, fake scarf/drug, fake glance/plan, false concert/war* and *former accusation/house*. We select these bigrams which are neither at ceiling or floor precisely because we suspect that there may be more than one context in participants' minds, and thus the specified contexts might split apart these middle-of-the-scale ratings into high (subsective) or low (privative) ratings respectively, explaining (some of) the variance in Experiment 2. Further, we select pairs of high and zero/low frequency bigrams because it is possible that high frequency bigrams might come with more conventionalised contexts and/or more fixed meanings and inferences in general, and thus might resist manipulation by provided contexts. Conversely, zero/low frequency bigrams might be particularly easy to manipulate, since they lack any preconceived "default" context. 40 native English speakers were recruited on Prolific (of which 1 excluded due to failed attention checks); the study was implemented in Qualtrics. Participants were paid pro rata at $12/hour; the experiment took 8 minutes on average. Each participant saw 12 items, as shown in Figure 7 (3 intersective-biased, 3 privative-biased, 6 fillers) yielding 10 ratings/item.

5.2. RESULTS. We find that across the board, writing biased contexts does indeed shift participants ratings in the intended direction. We fit an ordinal mixed effects model in R (R Core Team 2023, Christensen 2022) and find statistically significant effects for both the subsective-biased and privative-biased contexts compared to having no context ($p < 0.05$ for both; a subsective-biased context makes a high (subsective) rating 6x more likely while a privative biased context makes a high (subsective) rating only 0.5x as likely).

Looking at the data in detail, shown in Figure 8, we find that for some bigrams (*fake fire, fake plan, false concert*), the two contexts bias participants' entailments very effectively, moving the mean to ceiling or floor respectively and reducing the variance. We find a similar pattern for *fake*

scarf, fake glance and *counterfeit diamond*, which display the same movement towards the ends of the scale but with less reduction in variance. (For *counterfeit diamond* in particular, the population is split in half between definitely privative and definitely subsective, suggesting that the context only intervened as intended for half the participants.) Other bigrams have more mixed results. For *fake drug* and *false war*, we see an ability for the context to shift ratings to be more clearly subsective, but do not see the same for the supposedly privative-biased contexts. For *fake reef*, we still find that the same bigram can be either subsective or privative, since it is judged privative out of the blue in Experiment 2; however, the particular context supplied in this experiment did not result in privativity. We suspect the same issue for *former accusation* and *former house*, attributing the lack of change in ratings to experimenter misjudgement in writing the contexts: for example, for *former house*, two contexts were used where the house was converted into a hotel vs. became part of a museum; the former was not deemed privative and a context describing the demolition of the house would likely have been more successful. Finally, *counterfeit dollar* refuses to be influenced by context at all. This is likely nonetheless not an effect of conventionalization/memorization of its meaning or entailment, but rather that *dollars* depend so heavily on having an authentic method of manufacture that any possible way in which they can be counterfeited, i.e. in which their method of manufacture is non-conventional, robs them of being a dollar.

Turning to our second question, we find no effect of frequency in this experiment: high-frequency bigrams do not have more fixed entailments and are not more resistant to being manipulated by context. This might manifest in the form of high variance for the one context that goes against the bigram's "default" meaning or context, or in low variance ratings at the same end of the scale (whichever end is conventionalized) for both contexts. We see neither of these patterns.

5.3. DISCUSSION. The results from Experiment 3 show that the subsective/privative entailments drawn from adjective-noun combination are indeed context-dependent: it is possible to cause participants to draw completely subsective or completely privative inferences for the very same adjective-noun bigram by providing different contexts. This is possible whether the bigram is high-frequency (thus potentially coming with a "default" or conventionalized context of use) or novel. Thus it is likely that the participants' imagined contexts explains some or much of the variance in Experiment 2, which presented the bigrams out of the blue. While not all contexts in Experiment 3 fully achieve the desired shift in ratings, the results show that such manipulation is very much possible and likely could be achieved successfully for most bigrams (though not all – see *counterfeit dollar*). In other words, the meaning of adjective-noun bigrams and their privative entailments cannot be explained by memorization of a single (conventionalized) meaning or entailment, since entailments must be computed productively on the fly based on the provided context (as well as world knowledge). Finally, the ability to manipulate the entailments of novel bigrams such as *false concert* supports a compositional account for the meaning of the bigram (where context is included as part of the composition and inference-drawing process), just as we found in Experiment 2.

6. Summary of experiments. Our experiments showcase the wide variation in privative entailments among so-called privative adjectives: first, adjectives can license either a subsective or privative entailment depending on the noun (and context), and second, the same adjective-noun bigram can license either entailment depending on context. (Third, a given adjective-noun-context triple

may still receive graded/intermediate ratings, though this may be explained either by its entailment actually being intermediate/graded or by the context being insufficiently specified.) This poses challenges for theories which treat privativity as a property of the adjective (Partee 2010, del Pinal 2015, Guerrini 2022a), as we will discuss below. Further, this same variation is possible with novel adjective-noun bigrams: novel bigrams license both subsective and privative entailments, and their variance in ratings is similar to high-frequency bigrams; we do not find any effects of frequency/convention in either Experiment 2 or Experiment 3. Thus, despite the high degree of variation between and within adjectives and bigrams, these results support a compositional, context-dependent account of adjective-noun modification rather than an approach based on prior experience or convention which memorizes the meanings and/or entailments for previously encountered bigrams.

7. Theoretical accounts. A classic account for privative adjectives is due to Partee (Kamp & Partee 1995, Partee 2007, 2009, 2010). This account improves over the set-complement approach mentioned in the introduction (that $[\![\text{fake gun}]\!] \subseteq U \setminus [\![\text{gun}]\!]$) by positing that all seemingly privative adjectives in fact compose subsectively with the noun (i.e. as a function which creates a subset of the noun), like all other adjectives. The default interpretation of the noun is still literal, such that e.g. *fire* includes only real fires. So (unlike regular subsective adjectives) $[\![\text{fake fire}]\!] = \varnothing$ by default. To get a non-empty denotation, Kamp & Partee propose the Non-Vacuity Principle, which *fake fire* violates by having an empty extension. This violation is repaired by coercing *fire* to expand to include both real and fake fires, such that *fake* can now act subsectively over this new expanded set.

(10) **Non-Vacuity Principle**
In any given context, try to interpret any predicate so that both its positive and negative extension are non-empty (Kamp & Partee 1995; p. 161)

While Partee's approach does a good job of capturing the broad strokes of privative adjective behaviour, in particular the idea that they are just functions over the noun like any other adjective, it is somewhat underspecified, and issues remain with the particular implementation. For one, this still fixes privativity as a property of adjectives: *fake* necessarily yields things which are not in the literal denotation of *fire*, and then gets rescued by the Non-Vacuity Principle. This is difficult to reconcile with our experimental data that a *fake fire* may very well be a *fire*. Martin (2022) also shows experimentally that the default interpretation of a noun without any adjective need not be literal, and can include non-real instances such as toys (e.g. teddy bears are selected when participants are asked for *bears*), even if words like *toy* or *fake* are never explicitly mentioned. While one might try to salvage this by arguing that the second use of *fire* or these broad uses of *bear* also undergo coercion to a larger set of fires or bears, it is unclear what would trigger this coercion, since the Non-Vacuity Principle applies at predicate level and *fire* or *bear* alone meet their requirements just fine. The coercion would have to be at a pragmatic level after the whole sentence is computed or after the toy-bear-containing context is accounted for. Secondly, several responses to Partee (del Pinal 2015, 2018, Guerrini 2022a) argue that this coerced expansion is highly underspecified in terms of what it expands to include, and that this account does not actually

provide an explicit semantics for *fake*. As specified, it overgenerates, such that (11) could in principle turn out to be true (if you can expand to include any non-gun).

(11)   That giraffe is a fake gun.

Presumably, the expansion is intended to be context-dependent, such that the context dictates what can and can't be included (and excludes giraffes), but this is not spelled out in Partee's work. Further, more plausible scenarios such as *fake toy guns* which happen to be actual *guns*, not *toy guns* (but look like toy guns, thus *fake toy guns*), are not handled well (del Pinal 2015).

In the remainder of this section, we will discuss two types of accounts which are compositional while still accounting for the context-dependence (including noun-dependence) of privative entailments shown by the above experiments. Other accounts of privative adjectives based on cognitive or conceptual semantics, such as as shifting senses or conceptual blending (Franks 1995, Coulson & Fauconnier 1999), have also been proposed, but we will restrict ourselves here to proposals which have been implemented in formal semantics.

7.1. QUALIA-BASED ACCOUNTS. Del Pinal (2015, 2018) argues that we can have a compositional, truth-conditional account and still explain the behaviour of adjectives like *real, typical* and *fake* in a principled manner by moving to a two-dimensional semantics. Del Pinal's Dual Content Semantics proposes that we enrich our semantics to have an extensional component (E-structure), which corresponds to the set meanings that we are used to, plus a conceptual component (C-structure). This C-structure essentially captures the concept behind the noun, leaning on the large body of literature on concepts in psychology to do so. This C-structure is structured into qualia. Del Pinal uses the four qualia in Pustejovsky (1991) (CONSTITUTIVE, FORMAL, TELIC and AGEN-TIVE), though he is not strictly committed to them. The intent is that "[the] C-structure of lexical items encodes what, according to our best psychological theories, are the basic components of the corresponding kinds of concepts" (del Pinal 2018). Importantly, the E-structure and C-structure of a noun are separate, and neither one is derived from the other. This accounts for the fact that what actually counts as a *fire* or *lion* (what is in its extension) is somewhat independent of the prototypical description we have in our minds for *lion*, just as when we make statements about kinds such as *lions have manes* or *lions have four paws*, despite the existence of female lions and individual three-pawed lions which still qualify as lions. (See del Pinal (2015) for further justification and discussion of issues with deriving E-structure from C-structure.)

The C-structure of a noun comes into play for compositional purposes when we meet adjectives like *fake*, which not only update the C-structure of the resulting combination (as all adjectives do) but also use the C-structure of e.g. *gun* to build the E-structure for *fake gun*. In del Pinal's example, *gun* has the semantics shown in (12): guns are made of gun parts, look like guns, have the goal of being instruments of shooting events, and are made with that goal in mind.

(12) $\llbracket$gun$\rrbracket$ =   (del Pinal 2015; p.11)
        E-structure: $\lambda x.\, \mathrm{gun}(x)$

14

C-structure:

CONSTITUTIVE:	$\lambda x. \text{parts-gun}(x)$
FORMAL:	$\lambda x. \text{perceptual-gun}(x)$
TELIC:	$\lambda x. \text{GEN}\, e[\text{shooting}(e) \wedge \text{instrument}(e, x)]$
AGENTIVE:	$\lambda x. \exists e_1\, [\text{making}(e_1) \wedge \text{goal}(e_1, \text{GEN}\, e[\text{shooting}(e) \wedge \text{instrument}(e, x)])]$

Then, *fake* modifies *gun* to yield the semantics in (13) for *fake gun*: firstly, in the E-structure, *fake guns* are not in the extension of guns. Secondly, *fake guns* do not have the origins of guns (the agentive property), instead, they were made to appear as if they were guns. Moving to the C-structure, this is also the new agentive property, and means that they have the appearance of guns (the same formal properties) and do not have the same purpose (telic property). Formally, this is implemented by *fake* using the operators Q_E, Q_C, Q_A, Q_F and Q_T to access the extension and the four qualia of *gun* respectively (e.g. $\neg Q_A(\llbracket \text{gun} \rrbracket)$ negates the agentive quale of *gun*).[11] This is similar to Partee's proposal that privative adjectives are subsective like all other adjectives, in that all adjectives, including privative adjectives, operate in the same way on nouns (namely as functions over E-structure and C-structure), except we are not committed to the fact that the resulting E-structure is a *subset* of the E-structure of *guns*.

(13) $\llbracket \text{fake gun} \rrbracket =$ (del Pinal 2015; p.21)

E-structure: $\lambda x. \neg Q_E(\llbracket \text{gun} \rrbracket)(x) \wedge \neg Q_A(\llbracket \text{gun} \rrbracket)(x)$
$\wedge \exists e_2\, [\text{making}(e_2) \wedge \text{goal}(e_2, Q_F(\llbracket \text{gun} \rrbracket)(x))]$

C-structure:

CONSTITUTIVE:	$Q_C(\llbracket \text{gun} \rrbracket) = \lambda x. \text{parts-gun}(x)$
FORMAL:	$Q_F(\llbracket \text{gun} \rrbracket) = \lambda x. \text{perceptual-gun}(x)$
TELIC:	$\neg Q_T(\llbracket \text{gun} \rrbracket) = \lambda x. \neg \text{GEN}\, e[\text{shooting}(e) \wedge \text{instrument}(e, x)]$
AGENTIVE:	$\lambda x. \exists e_2\, [\text{making}(e_2) \wedge \text{goal}(e_2, Q_F(\llbracket \text{gun} \rrbracket)(x))]$

While del Pinal's theory does an excellent job specifying how privative adjectives like *fake* interact with the noun and what the resulting concept looks like, as well as handling some interesting edge cases like *fake toy gun* or *fake red Ferrari*, del Pinal (2015) still stipulates, like Partee, that *fake guns* are not *guns* by fixing in the E-structure that a *fake gun* is not in the extension of *gun*: $\neg Q_E(\llbracket \text{gun} \rrbracket)(x)$. In subsequent work, del Pinal (2018) admits that this part of the E-structure is questionable for *counterfeit* and *artificial* and that it is an empirical question whether this should be included or not. Empirically, both we and Martin (2022) find that it should not, necessitating an adjustment to del Pinal's definitions.

---

[11]Martin (2019) and Guerrini (2022a) point out issues with this particular lexical entry for *fake* with regards to multiple application of C-structure-accessing adjectives (*fake fake gun* or *fake typical gun*). Martin (2019) proposes a tweak to the lexical entry to fix the problem, while Guerrini argues that even this fix does not make sufficient/correct predictions. We are not committed to the exact implementation of the lexical entry for *fake* or for *typical* here, as long as it can be expressed in del Pinal's formalism.

Adjusting del Pinal's definition to remove $\neg Q_E(\llbracket \text{gun} \rrbracket)(x)$ is exactly what Martin (2022) proposes.[12] The main ingredient to make privativity emergent rather than fixed is simply to remove the stipulation that a *fake gun* is not in the extension of *gun*, shown in (14).

(14) $\llbracket \text{fake gun} \rrbracket =$ (Martin 2022; p.207)
E-structure: $\lambda x. \neg Q_A(\llbracket \text{gun} \rrbracket)(x) \wedge \exists e_2 \, [\text{making}(e_2) \wedge \text{goal}(e_2, Q_F(\llbracket \text{gun} \rrbracket)(x))]$
C-structure: same as in (13)

If we do not include privativity in the definition of *fake*, when does privativity come about? Martin says that it occurs when the targeted, negated quale is "central" to the meaning of the noun, i.e. part of the E-structure. This captures the variation by noun, since even if an adjective targets the same qualia across nouns, these qualia may or may not be "central" to this noun. In fact, del Pinal (2018) also discusses centrality of qualia (implemented by structuring the qualia as an ordered tuple from most to least central), though not in the context of privativity, suggesting that certain adjectives (and possibly certain contexts?) may modify which qualia are most central to a noun. Martin also hints that context may affect how central a quale is to the noun meaning. Though Martin only discusses context in response to Guerrini (2018), who claims that context can affect which prototypical features/qualia the adjective targets (Martin argues that it should not, in part because this predicts again that *fake* is always privative), this is clearly the place to modify this theory to account for the effect of context and world knowledge that we see in Experiment 3. For now, we can simply say that context affects which qualia are central to noun membership (E-structure) similarly to how context plays a role in other semantic areas such as domain restriction or the availability of scalar alternatives, but we hope to lay this out in more detail in future work.[13] Under any formulation, this process is clearly productive and lets us assess bundles of qualia produced by novel adjective-noun bigrams in contexts that we have never encountered before, explaining the results for novel bigrams in Experiment 3.

In other words, under this account, composition is just the first step of privativity. Composition proceeds the same for all adjective-noun pairs, namely as as a function over noun qualia. A second step evaluates this new bundle of qualia for noun membership, by examining which qualia are still present and whether they are central to being an instance of that noun or not, to derive subsective/privative entailments. This second step accounts for the effect of context, which influences which qualia matter for determining noun membership.

7.2. SIMILARITY-BASED ACCOUNTS. Guerrini (2022a) proposes a second compositional account for privativity. Guerrini builds on the intuitions for the lexical entries of *fake* and *counterfeit* in del Pinal (2015), but argues that there is no need for a two-dimensional semantics. Instead, he proposes a similarity-based account which handles the meaning of *fake* using *seems like*, using the one-dimensional compositional semantics for *seems like* from Guerrini (2022b).

---

[12]Martin also proposes a modification to del Pinal's broader account, namely that the E-structure should be derived from the C-structure after all, and that in fact the C-structure should be replaced entirely with semantic kinds, and converted to an extension (to E-structure) in the syntax. Whether this move is necessary and whether it holds up to the criticisms that del Pinal discusses about theories which derive the E-structure from the C-structure is an open issue.

[13]One interesting avenue might be to link the context-dependent mod_l in del Pinal (2018), which provides the correct "implementation" of the noun given the context, with the centrality feature and with which qualia are included in the E-structure of contextually-modulated *gun* vs. the newly composed (and also contextually modulated) *fake gun*.

(15) $[\![\text{fake}]\!] = \lambda P \lambda x.\, \textsc{Intended}(x, \textsc{seem-like}(x, P)) \wedge \neg P(x)$ (Guerrini 2022a; p.17)

(16) $\textsc{Intended}(x, P) = \exists e.\, \textsc{Action}(e, x) \wedge \textsc{Goal}(e, P(x))$

This captures the same intuitions as del Pinal (2015), namely that *fake guns* are not guns but are built to seem like (compare: have the perceptual qualia of) guns. The difference is is that the purpose and origin (telic and agentive qualia) of the *fake gun* are not specified, to avoid paradoxes with combinations like *fake fake gun*. Note that like the original definition of *fake* in del Pinal (2015), and in keeping with Partee, privativity is explicitly baked into Guerrini's definition of *fake*. Unlike del Pinal, Guerrini is explicitly committed to this fact. He argues that the variability in privativity in the experiments in Martin (2022), which we have replicated and extended in this paper, is in fact explained by syntactic ambiguity. Specifically, in an earlier part of his dissertation, Martin (2022) proposes that adjective modification may be either direct modification (which is always subsective) or indirect modification, in which case the adjective modifies a covert noun (subsectively), similar to analyses of adjectives as reduced relative clauses. Martin proposes this in order to account for the ambiguity between readings of bigrams like *good thief*, which may have a subsective reading of *is good at thievery* or a truly intersective reading of *is (morally) good and a thief*, as well as puzzles involving kind readings and number in languages such as Turkish. Martin derives this intersective reading by arguing that the adjective is actually subsectively modifying some context-appropriate, null noun such as *person*, which is adjoined higher in the syntactic structure. This results in apparent intersectivity. (The null noun may also be identical to the overt one, yielding the same subsective reading as if the adjective were low, but with different implications for the other phenomena that Martin discusses.)

(17) Subsective reading
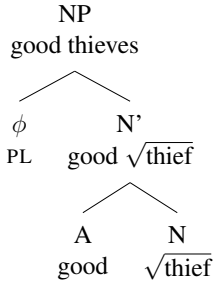
(18) Intersective reading
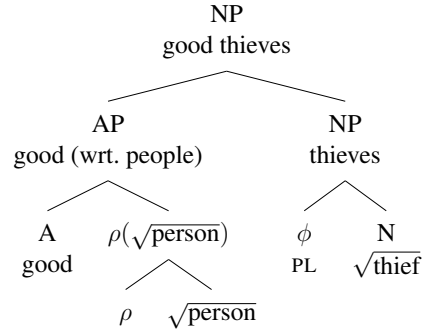
Guerrini takes this syntactic ambiguity account, in conjunction with some Italian data where the two readings come apart more clearly, to point out that when asking something like *Is a fake watch a watch?* in Martin's experiment, or in our Experiment 2, *fake watch* is ambiguous between (amongst other things) *[[fake wrt. Rolexes] watch]* and *[fake (wrt. watches) watch]*. Using this ambiguity, Guerrini can maintain that *fake* is always privative. He argues that in the former case, *fake* applies privatively to *Rolex*, not *watch*, resulting in something that is not a Rolex but is still a watch. In other words, every time participants get a subsective entailment in Experiment 2 or 3, it is because they resolve the syntactic ambiguity of *fake N* with indirect modification and

choose a covert noun which is not *N*. They get the privative entailment when they choose direct modification. This choice presumably depends on whether the context supplies an appropriate covert noun, allowing privativity to be context-dependent.

However, since Guerrini insists that *fake* is privative – i.e. that *fake guns* are not *guns* – his account faces one problem: why can we refer to *fake guns* as *guns*, as in (19)?

(19)   That gun is fake.

While Guerrini proposes a fix similar to Partee's Non-Vacuity Principle that works specifically for copulas involving *fake*, the broader question remains open (as it does for Partee) of how to handle references to fake or toy items that do not use *fake* explicitly, such as (7c) or (20). (Qualia-based accounts get around this by deferring the problem to whatever is in the E-structure, which is affected by context and may or may not include *fake* things which do not meet the standards of the noun's C-structure.)

(7c)   This [*pointing at a toy bear*] is a bear.

(20)   A: We have a wide selection of props, including a fake lightsaber and a fake gun.
       B: Pass me the gun.

Nevertheless, for the purposes of our experiments, Guerrini's account explains the variation that we see, assuming (as seems plausible) that he can provide adequate definitions of our other adjectives *artificial, counterfeit, false, former* and *knockoff* using his tools of SEEM-LIKE and intentions.

7.3. SUMMARY. While the results from our experiments cause difficulties for accounts like Partee (2010) which fix privativity as a property of the adjective and only have a single mechanism of adjective modification, the variation in privative entailments can be captured compositionally either by making privativity emergent (context-dependent) or by keeping privativity as a property of the adjective but appealing to ambiguity in which noun it is applied to. Both the modification by Martin (2022) of the qualia-based account in del Pinal (2015, 2018) and the account in Guerrini (2022a) thus capture all the variation we see in Experiments 2 and 3 (within-adjective between nouns, and within-bigram between contexts) while retaining a compositional theory. This explains why participants have no trouble understanding and drawing entailments for novel bigrams. Martin (2022) further predicts what kind of entailments they should draw when, provided that we understand the effect of the context on the central qualia for the noun.

8. Conclusion. This paper presents experimental evidence on the variation in subsective vs. privative entailments both within adjectives and within adjective-noun bigrams, including for novel adjective-noun bigrams. We find, as does Martin (2022), that no adjective always yields privative entailments, and further we find that no adjective-noun pair is fixed in its entailments either. Privativity depends on the adjective, the noun and the context. Nevertheless, we find support for a compositional account of this process, provided that context is included as part of composition. One might believe that broad set of effects of applying adjectives like *fake* to a noun and the resulting variation in entailment might be so broad as to need to be memorized per-context, or per-bigram, such that we have fixed meanings or entailments (perhaps with accompanying default contexts)

18

for high-frequency, well-known bigrams like *fake fire, artificial light* or *false information* and do not assign a particularly clear meaning or entailment to other bigrams until we have experienced them (enough times) in context. We find no such effect: not only do participants interpret novel bigrams successfully and with a similar overall variance in ratings as high frequency bigrams, but the entailments of high frequency and novel bigrams can both be heavily influenced by context in either direction. This suggests that the process of building a new concept from an adjective modifying a noun is compositional and the process of deriving the subsective or privative entailment is productive as a function of adjective and noun meaning (including world knowledge) plus context. Specifically, these results supports compositional accounts like Martin (2022)'s modification of del Pinal (2015, 2018) and Guerrini (2022a) which predict that privativity is context-dependent, and raises issues for theories which are not compositional (e.g. basic set complementation) or which fix privativity as a property of individual adjectives (Partee 2010, del Pinal 2015) with only a single method of composition.

References

Arnon, Inbal & Neal Snider. 2010. More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language* 62(1). 67–82. 10.1016/j.jml.2009.09.005. https://www.sciencedirect.com/science/article/pii/S0749596X09000965.

Caldwell-Harris, Catherine, Jonathan Berant & Shimon Edelman. 2012. Measuring Mental Entrenchment of Phrases with Perceptual Identification, Familiarity Ratings, and Corpus Frequency Statistics. In *Measuring Mental Entrenchment of Phrases with Perceptual Identification, Familiarity Ratings, and Corpus Frequency Statistics*, 165–194. De Gruyter Mouton. https://www.degruyter.com/document/doi/10.1515/9783110274073.165/html.

Christensen, R. H. B. 2022. ordinal—Regression Models for Ordinal Data. R package version 2022.11-16. https://CRAN.R-project.org/package=ordinal.

Coppock, Elizabeth & Lucas Champollion. 2023. Invitation to Formal Semantics. https://eecoppock.info/semantics-boot-camp.pdf.

Coulson, Seana & Gilles Fauconnier. 1999. Fake guns and stone lions: Conceptual blending and privative adjectives. In Barbara A. Fox, Dan Jurafsky & Laura A. Michaelis (eds.), *Cognition and function in language*, 143–158. Center for the Study of Language and Information.

Dodge, Jesse, Maarten Sap, Ana Marasović, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell & Matt Gardner. 2021. Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus. 10.48550/arXiv.2104.08758. http://arxiv.org/abs/2104.08758.

Franks, Bradley. 1995. Sense Generation: A "Quasi-Classical" Approach to Concepts and Concept Combination. *Cognitive science* 19(4). 441–505. 10.1207/s15516709cog1904$_2$.

Guerrini, Janek. 2018. The link between misinterpretation, intentionality, and mental agency in the natural language interpretation of "fake". *Rivista Italiana di Filosofia Analitica Junior* 9(2). 181–192. https://riviste.unimi.it/index.php/rifanalitica/article/view/11095.

Guerrini, Janek. 2022a. Keeping fake simple. Accepted with minor revisions in Journal of Seman-

tics. https://lingbuzz.net/lingbuzz/006977/current.pdf.

Guerrini, Janek. 2022b. 'Like a N' constructions: genericity in similarity. In Dean McHugh & Alexandra Mayn (eds.), *Proceedings of the ESSLLI 2022 Student Session*, Galway. https://doi.org/10.21942/uva.20368104. https://doi.org/10.21942/uva.20368104.

Heim, Irene. 1991. Artikel und Definitheit (Articles and Definiteness). In *Semantik / Semantics*, vol. 6, 487–535. Berlin, New York: De Gruyter Mouton republished 2010 edn.

Heim, Irene & Angelika Kratzer. 1998. *Semantics in generative grammar* Blackwell textbooks in linguistics 13. Malden, Mass., USA: Blackwell.

Kamp, Hans & Barbara H. Partee. 1995. Prototype theory and compositionality. *Cognition* 57(2). 129–191. 10.1016/0010-0277(94)00659-9.

Martin, Joshua. 2019. Compositionality in Privative Adjectives: Extending Dual Content Semantics. In Jennifer Sikos & Eric Pacuit (eds.), *At the Intersection of Language, Logic, and Information* Lecture Notes in Computer Science, 93–107. Berlin, Heidelberg: Springer. 10.1007/978-3-662-59620-3$_6$.

Martin, Joshua. 2022. *Compositional Routes to (Non)Intersectivity*. United States – Massachusetts: Harvard University Ph.D. https://www.proquest.com/docview/2681380157/abstract/58A63B8C3E6548AEPQ/1.

Nayak, Neha, Mark Kowarsky, Gabor Angeli & Christopher D. Manning. 2014. A Dictionary of Nonsubsective Adjectives. Tech. Rep. CSTR 2014-04 Department of Computer Science, Stanford University. https://www-cs.stanford.edu/~angeli/papers/2014-tr-adjectives.pdf.

O'Donnell, Timothy J. 2015. *Productivity and Reuse in Language: A Theory of Linguistic Computation and Storage*. Cambridge, Massachusetts, London, England: The MIT Press. https://doi.org/10.7551/mitpress/9780262028844.001.0001.

Partee, Barbara H. 2007. Compositionality and coercion in semantics: The dynamics of adjective meaning. *Cognitive foundations of interpretation* 145–161.

Partee, Barbara H. 2009. Formal semantics, lexical semantics, and compositionality: The puzzle of privative adjectives. *Philologia* 7(1). 11–21. http://www.philologia.org.rs/index.php/ph/article/view/216.

Partee, Barbara H. 2010. Privative adjectives: Subsective plus coercion. In Thomas Zimmermann, Rainer Bauerle & Uwe Reyle (eds.), *Presuppositions and discourse: Essays offered to Hans Kamp*, 273–285. Brill. https://brill.com/downloadpdf/book/edcoll/9789004253162/B9789004253162-s011.pdf.

Pavlick, Ellie & Chris Callison-Burch. 2016a. Most "babies" are "little" and most "problems" are "huge": Compositional Entailment in Adjective-Nouns. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2164–2173. Berlin, Germany: Association for Computational Linguistics. 10.18653/v1/P16-1204. https://aclanthology.org/P16-1204.

Pavlick, Ellie & Chris Callison-Burch. 2016b. So-Called Non-Subsective Adjectives. In *Proceedings of the Fifth Joint Conference on Lexical and Computational Semantics*, 114–119. Berlin, Germany: Association for Computational Linguistics. 10.18653/v1/S16-2014.

https://aclanthology.org/S16-2014.

del Pinal, Guillermo. 2015. Dual Content Semantics, privative adjectives, and dynamic compositionality. *Semantics and Pragmatics* 8. 7:1–53. 10.3765/sp.8.7. https://semprag.org/index.php/sp/article/view/sp.8.7.

del Pinal, Guillermo. 2018. Meaning, modulation, and context: a multidimensional semantics for truth-conditional pragmatics. *Linguistics and Philosophy* 41(2). 165–207. 10.1007/s10988-017-9221-z. https://doi.org/10.1007/s10988-017-9221-z.

Pustejovsky, James. 1991. The Generative Lexicon. *Computational linguistics - Association for Computational Linguistics* 17(4). 409–441.

Pustejovsky, James. 2013. Inference Patterns with Intensional Adjectives. In *Proceedings of the 9th Joint ISO - ACL SIGSEM Workshop on Interoperable Semantic Annotation*, 85–89. Potsdam, Germany: Association for Computational Linguistics. https://aclanthology.org/W13-0509.

R Core Team. 2023. R: A Language and Environment for Statistical Computing. https://www.R-project.org/.

Raffel, Colin, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li & Peter J. Liu. 2020. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research* 21(140). 1–67. http://jmlr.org/papers/v21/20-074.html.

Szabó, Zoltán Gendler. 2012. The case for compositionality. In Wolfram Hinzen, Edouard Machery & Markus Werning (eds.), *The Oxford Handbook of Compositionality*, Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199541072.013.0003.

Tremblay, Antoine & Harald Baayen. 2010. Holistic Processing of Regular Four-word Sequences: A Behavioural and ERP Study of the Effects of Structure, Frequency, and Probability on Immediate Free Recall. *Perspectives on Formulaic Language: Acquisition and Communication* 151.

Vecchi, Eva M., Marco Marelli, Roberto Zamparelli & Marco Baroni. 2017. Spicy Adjectives and Nominal Donkeys: Capturing Semantic Deviance Using Compositionality in Distributional Spaces. *Cognitive Science* 41(1). 102–136. 10.1111/cogs.12330. https://onlinelibrary.wiley.com/doi/abs/10.1111/cogs.12330.

Zehr, Jeremy & Florian Schwarz. 2018. PennController for Internet Based Experiments (IBEX). https://doi.org/10.17605/OSF.IO/MD832.