Polysemy and Inference: Reasoning with Underspecified Representations

Elliot Schwartz[1,2,*], **Griffin Pion**[1,2,*], **Jake Quilty-Dunn**[5], **Eric Mandelbaum**[1,2,3,6], **Spencer Caplan**[1,4]

eschwartz@gradcenter.cuny.edu, gpion@gradcenter.cuny.edu, jake.quilty-dunn@rutgers.edu, eric.mandelbaum@baruch.cuny.edu, scaplan@gc.cuny.edu

[1]Cognitive Science, [2]Philosophy, [3]Psychology, [4]Linguistics Programs, CUNY Graduate Center, USA
[5]Department of Philosophy and Center for Cognitive Science, Rutgers University, USA
[6]Department of Philosophy, Baruch College, USA

Abstract

Lexical ambiguity has classically been categorized into two kinds. Homonyms are single word forms that map to multiple, unrelated meanings (e.g., "bat" meaning baseball equipment or a flying mammal). Polysemes are single word forms that map to multiple, related senses (e.g., "breakfast" meaning a plate of food or an event). Yet there is a longstanding debate as to whether polysemy and homonymy reflect distinct cognitive representations. Some (e.g., Fodor & Lepore, 2002; Klein & Murphy, 2001) posit that they do not — merely describing differing patterns of usage — while others (e.g., Falkum & Vicente, 2015; Pietroski, 2018) argue that polysemes, but not homonyms, involve an underspecified representation that is neutral with respect to the form's multiple senses. While some extant experimental evidence supports the latter view (Klepousniotou, Titone, & Romero, 2008; Srinivasan, Berner, & Rabagliati, 2019), there has not yet been clear evidence of the representation of lexical ambiguity affecting domain-general reasoning. Using a novel inference paradigm, we compare participants' dispositions to endorse deductive, Aristotelian arguments with equivocating polysemes versus comparable arguments with equivocating homonyms. We find that participants endorse the former substantially more than the latter, a phenomenon that we dub the *Uncommon Sense Effect*. Our results provide direct evidence that polysemes and homonyms have underlyingly distinct mental representations — in particular that polysemes uniquely invoke an underspecified representation that allows for rule-based inferences across distinct senses.

Keywords: polysemy; homonymy; ambiguity; mental representation; lexical processing; inference; concepts; deduction

Introduction

Lexical ambiguity has classically been categorized into two kinds (e.g., Frazier & Rayner, 1990; Lyons, 1977; Nida, 1979, inter alia). *Homonyms* are single word forms that map to multiple, unrelated meanings. For example, "bat" refers to both flying mammals and clubs used in baseball. *Polysemes*, in contrast, are single word forms that map to multiple, related senses. For instance, "breakfast" refers to both dining events and foods. While related, these senses of "breakfast" are not identical: breakfast *foods* are the sort of things that can be overcooked, and breakfast *events* are the sort of thing that can last a short time, but *not vice versa*. There is mounting evidence that polysemes, unlike homonyms, follow systematic and cross-linguistic patterns. For example, Srinivasan and Rabagliati (2015) found that many patterns of polysemy appear to be typological universals. At least across the 15 languages tested, each language evidenced using the same word for an animal and its meat (e.g., "chicken") and the same word for a material and an artifact characteristically made of that material (e.g., "glass"). In contrast, the cross-linguistic distributions of homophonous meanings are arbitrary. Nevertheless, the cognitive representations supporting such typological patterns remain uncertain.

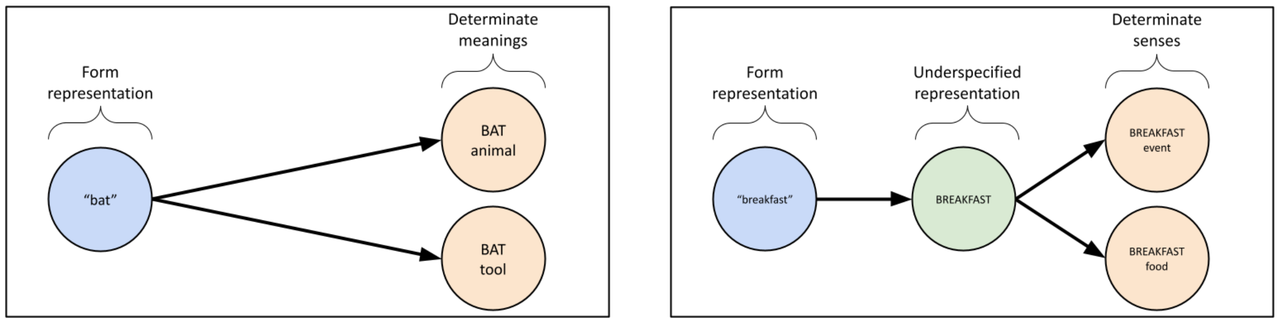
There is general consensus (Frisson, 2015) that homonyms employ a *one-to-many* mapping between a single word form ("bat") and two or more distinct senses which are stored separately (BAT_{animal} and BAT_{tool} — see Fig. 1a). However, accounts disagree regarding the status of polysemy. One account, which we term the **List View**, argues that polysemes employ the same representational structure as homonyms (Devitt, 2021; Fodor & Lepore, 2002; Klein & Murphy, 2001), with the difference between "classes" of ambiguity reduced to relatedness (either semantic or metalinguistic) between senses.[1] A different account, which we term the **Underspecification View**, holds that polysemes, unlike homonyms, involve a *one-to-one-to-many* structure: a single word form (e.g., "breakfast") maps to a single underspecified representation ($\text{BREAKFAST}_{underspecified}$), and this underspecified representation in turn points to multiple determinate senses (BREAKFAST_{event} and BREAKFAST_{food}). See Frisson (2009, 2015); for a review, Falkum and Vicente (2015); Fig. 1b).[2] If polysemes do involve an underspecified representation, there is a further question as to when this representation is activated during lexical processing. We return to this point in the Discussion.

In this paper, we develop a novel inference experiment on which the List and the Underspecification Views make divergent predictions. We thus investigate whether polysemes and homonyms have distinct mental representations. A number of other paradigms have been used to probe for processing differences between homonymy and polysemy including sensicality judgments (Frisson, 2015; Klein & Murphy, 2001; Kle-

*ES and GP are co-first authors

[1]For the view that polysemes are distinct from homonyms merely in the degree of their semantic representations' relatedness, see Devitt (2021). For the view that polysemes are distinct from homonyms in that only the former involve metalinguistic beliefs about relatedness, see Fodor and Lepore (2002)

[2]We include under the label "Underspecification View" any theory that posits an *intermediate representation* that is common across distinct senses, regardless of whether this representation is semantically sparser or richer than the specific senses cf. Falkum and Vicente (2015, p. 6-7).



(a) (b)

Figure 1: Two forms of lexical ambiguity. On the **List View**, both homonyms and polysemes are represented as a single word form mapping directly to its two or more constituent senses (panel a). On the **Underspecification View**, homonyms are represented as on the List View (panel a), but for polysemes, a single word form maps to an intermediate, underspecified structure which, in turn, maps to the word form's multiple determinate senses (panel b).

pousniotou et al., 2008), lexical decision tasks (Armstrong & Plaut, 2008, 2016; Beretta, Fiorentino, & Poeppel, 2005; Klepousniotou & Baum, 2007; Rodd, Gaskell, & Marslen-Wilson, 2002), eye-tracked reading (Frazier & Rayner, 1990; Frisson & Pickering, 1999, 2007; McElree, Frisson, & Pickering, 2006), as well as word learning and meaning extension (Rodd et al., 2012; Srinivasan & Snedeker, 2011). While these studies have provided insight into the potential processing differences and usage choices between polysemy and homonymy, such tasks do not clearly identify the representational status of such differences. Importantly, these methods are largely silent on whether polysemy is limited to linguistic processing (Pietroski, 2018) or whether it reflects underlying conceptual structure (Quilty-Dunn, 2021; Srinivasan & Rabagliati, 2015).

To this end, we developed a novel paradigm involving inference. Arguments are invalid when they equivocate on the meaning of a term. For example, although it is true that Anne Hathaway was Shakespeare's wife[3] and that Anne Hathaway has won an Academy Award, it would be wrong to infer that Shakespeare's wife won an Academy Award. Arguments that equivocate using different senses of a polyseme are likewise invalid. While it is true that France, the nation, has won the World Cup and that France, the geographic region, is a hexagon, it doesn't follow that a hexagon won the World Cup. We reason that since concepts are characterized by their involvement in rule-based inference (Carey (2009) p. 487; G. Murphy (2002); Peacocke (1992)), an inference task offers a straightforward domain to distinguish the List and Underspecification Views.

Expanding on this intuition, since the List View posits that both homonyms and polysemes involve the same representational structure (i.e., a direct one-to-many mapping from word forms to meanings), it predicts arguments with equivocal polysemes and homonyms should be judged invalid at similar rates. Furthermore, this propensity to reject these "equivocating" arguments should be quite high: nothing connects the two meanings involved aside from a shared word form. Conversely, since the Underspecification View posits that, specifically for polysemes, there is an *intermediate* representation invoked *before* resolution into any determinate sense, it predicts arguments with equivocal polysemes should be judged to be valid significantly **more** than comparable arguments involving equivocal homonyms. We call this hypothesized boost in perceived validity for equivocal polysemes the *Uncommon Sense Effect*. It would be the intermediate, underspecified representation accessed across polysemous premises which supports the Uncommon Sense Effect, a structure that the homonyms lack.

The paper is structured in four sections. First, we describe the construction of Aristotelian syllogisms which serve as our stimuli. These consist of argument forms containing either polysemous or homonymous words across premises, along with non-ambiguous filler items that are strictly valid or invalid for comparison. Next, in Experiment 1, we evaluate the distribution of judgments made over just the conclusions of the arguments. This allows us to fully equate the baseline believability of the conclusions for polysemous and homonymous stimuli; thus, any subsequent differences between the arguments involving the two groups of stimuli must result from the properties of the argument as a whole. Then, in Experiment 2, we evaluate the predictions of the List and Underspecification Views. We find evidence for the Uncommon Sense Effect: participants endorse the arguments with equivocal polysemes more often than they endorse comparable arguments with equivocal homonyms. This finding is consistent with the Underspecification View but not with the List View. Finally, we discuss these results, connect them with other research on polysemy, and draw out broader implications.

Argument Construction

We selected four Aristotelian syllogisms to serve as argument forms. Each syllogism consists of two premises sharing a

[3] Yes, that was actually her name.

middle term (M) along with a conclusion. We selected specific syllogisms based on two considerations. First, we used arguments that contained only affirmative statements (e.g., All M are P) as these are generally quicker and easier to process than negated statements (e.g., No M are P — Hasson and Glucksberg (2006); Reverberi, Pischedda, Burigo, and Cherubini (2012)). Second, we chose arguments with existentially quantified conclusions (e.g., Some M are P), since universal conclusions (e.g., All M are P) are rendered invalid by a single counter-instance, and are thus less plausible in general. This left us with the four argument forms referred to by medieval logicians as Dimatis, Disamis, Datisi and Darii (Table 1).

Dimatis	Disamis	Datisi	Darii
Sm P are M All M are S Sm S are P	Sm M are P All M are S Sm S are P	All M are P Sm M are S Sm S are P	All M are P Sm S are M Sm S are P

Table 1: The four Aristotelian syllogisms used in the present experiments.

We constructed arguments using polysemes and homonyms, as well as valid and invalid fillers. Polysemy arguments contained polysemes as the middle term (M) of an argument form. Each premise was written to uniquely constrain the sense of the polyseme being used. For example, in an argument using the polyseme "breakfast," the first premise concerned the event sense of "breakfast" (*"All breakfasts are under three hours"*), while the second concerned the food sense of "breakfast"(*"Some smoothies are breakfasts"*). Constructing the premises in this way uniquely determines a conclusion given the argument form (*"Some smoothies are under three hours"*). Homonymy arguments contained homonyms as the middle term (M) of an argument form. As with polysemy arguments, each premise was written to uniquely constrain the meaning of the homonym being used. For instance, in an argument using "bats", the first premise clearly concerned the animal meaning of "bat" (*"All bats share a common ancestor"*), while the second used the baseball meaning of "bat" (*"Some sporting goods are bats"*). Again, the conclusion (*"Some sporting goods share a common ancestor"*) is uniquely determined.

Valid and invalid fillers used middle terms (M) that did not contain lexical ambiguity and thus did not equivocate across the argument's premises. Invalid arguments were generated by taking each of the valid argument forms and swapping the order of terms in their universal premises; doing so results in an invalid argument for all four forms even though all premises were written to be independently plausible on their own.[4] Valid fillers were constructed by faithfully applying the same syllogistic forms but maintaining a valid argument by again eschewing lexical ambiguity.[5]

The twelve polysemous middle terms were drawn from E. Murphy (2021). These twelve terms comprised three widely discussed types of polysemy: *book-type* ("book", "magazine", "letter", "novel"), ambiguous between physical objects and information structures ("paperback novel" vs. "romance novel"); *lunch-type* ("lunch", "dinner", "picnic", "breakfast"), ambiguous between physical foods and events ("overcooked breakfast" vs. "short breakfast"); and *city-type* ("city", "company", "school", "college"), ambiguous between physical buildings and institutions ("dilapidated school" vs. "prestigious school"). We selected multiple types of polysemes to ensure that any effect is due to the general representation of polysemes, not to a feature of a particular type (e.g., *book-type*). For each polyseme, E. Murphy (2021, p. 96) also reported measures of sense dominance obtained from a cloze task. Since some work (e.g., Klepousniotou et al., 2008) suggests that the order of dominant and subordinate senses affects judgments, all polysemy arguments used the subordinate sense in the first premise and the dominant sense in the second. The majority of homonyms ("board", "suit", "cabinet", "bat", "ball", "ruler", "ring", "match", "star", "change") were drawn from Frazier and Rayner (1990) — with two additional words ("wave" and "rock") added to balance the number homonyms and polysemes.

Experiment 1: Conclusion Believability

A final constraint on stimulus construction was the believability of the arguments' conclusions. In research on inference, participants often exhibit a "belief bias" and judge arguments as valid more often if their conclusions are more believable (Evans, Barston, & Pollard, 1983). To control for this bias, we constructed an initially larger set of arguments and conducted a experiment in which participants judged the truth of each argument's conclusion in isolation — that is, without seeing the argument's premises at all. The stimuli for the main study (Experiment 2) could then be down-sampled from this set so that the average truth ratings for the conclusions of polysemous arguments *exactly* matched those for homonymous arguments.

Participants

Forty participants were recruited from Prolific and paid $2.25 for their participation. All participants were native English speakers residing in the United States who had completed at least 15 previous submissions on Prolific.

Materials

We evenly divided the twelve selected polysemes and twelve selected homonyms between Dimatis, Disamis, Datisi and Darii so that each syllogistic form was associated with three polysemes and three homonyms. For each of the word-argument pairs, we constructed either two or three argument

[4] An example Invalid Filler argument proceeds as follows: All medieval manuscripts are antiques. Some pieces of jewelry are antiques. ∴Some pieces of jewelry are medieval manuscripts.

[5] An example Valid Filler argument proceeds as follows: All dolls are man made. Some toys are dolls. ∴Some toys are man made.

candidates[6] — with the end goal of selecting a single candidate per word-argument pair to include in the main study.

This procedure yielded a total set of 88 arguments: 56 ambiguous target arguments (25 polysemy and 31 homonymy) along with 16 valid fillers and 16 invalid fillers evenly divided among the four argument types. Each participant was assigned to one of two randomized orderings of the items and provided judgments for all 88 conclusions.

Procedure

The experiment was conducted online using using PCIbex (Zehr & Schwarz, 2018). Participants were told they would see a series of statements and instructed to click a button marked "True" if the statement seemed true and a button marked "False" if the statement seemed false. After viewing examples of true and false statements, participants gave judgments on all 88 conclusions arranged in one of the two pre-randomized orders.

Results

Results from Experiment 1 indicate the success of our argument construction. The conclusions of the Valid Fillers were judged to be true nearly all of the time (i.e., ceiling performance of 94%). Correspondingly, the Invalid Filler conclusions were judged to be false a majority of the time: 28% average truth judgment. We used the conclusion judgments from Experiment 1 to first generate all possible combinations of the variants for each word-argument pair (22,057,920 in total). Among those, we identified the Polysemy/Homonymy combinations which had exactly equal mean truth judgments (there are 830 such sets) and manually selected the final stimulus set for use in Experiment 2 from among them. This set of 54 arguments (12 polysemy, 12 homonymy, 16 valid fillers, 16 invalid fillers) crucially has precisely the same mean truth rating (31%) for polysemes/homonyms — see Fig. 2 for the full distribution. Thus, any increase in acceptability of polysemy arguments is due to the additional context provided by the premises rather than the independent believability of the conclusion. Moreover, we established that on average the conclusions of the homonymous and polysemous items are equally as poorly rated as the conclusions of the invalid fillers (unpaired t-test: $p = 0.74$).

Experiment 2: Uncommon Sense Effect

To evaluate the differing predictions of the List View and the Underspecification View, we had participants judge the validity of arguments involving four word-types (polysemy, homonymy, along with valid and invalid fillers) evenly split among four Aristotelian argument forms (Dimatis, Disamis,

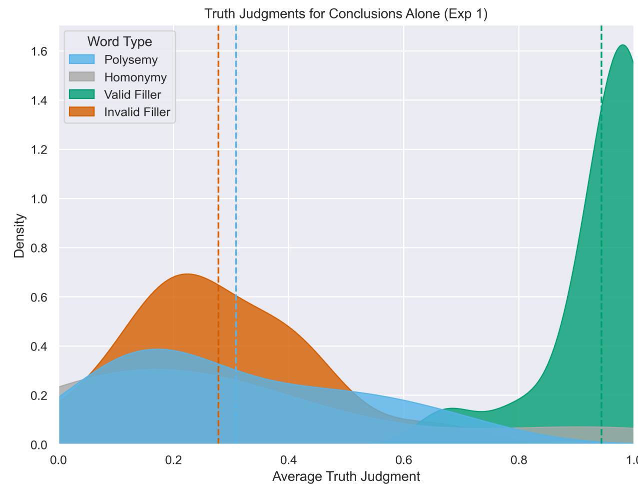


Figure 2: Smoothed distribution (KDE) of selected conclusion truth-judgments as a function of word-type. The Polysemy and Homonymy items were chosen to have equal means (0.31). Valid Filler: 0.94, Invalid Filler: 0.28.

Datisi and Darii). These were fully crossed resulting in a 4 (word-type) x 4 (argument form) within-subject design.

Participants

Twenty-eight participants were recruited from Prolific and paid $3.75 for their participation. All participants were native English speakers residing in the United States who had completed at least 15 previous submissions on Prolific and did not participate in Experiment 1.

Materials

Stimuli were the 56 arguments (12 polysemy, 12 homonymy, 16 valid fillers, 16 invalid fillers). Each participant provided validity judgments for all arguments ordered according to one of the two pre-randomized lists.

Procedure

As in Experiment 1, this experiment was conducted online using PCIbex (Zehr & Schwarz, 2018). Participants were told to assume that the premises of each argument were true and then to judge whether the conclusion followed. This instruction was intended to ensure that participants focused on the argument's validity, not the isolated truth of the premises or conclusion. We then provided them with some examples of valid and invalid syllogisms. For the experimental trials, only one of three statements in the argument (premise 1, premise 2, or the conclusion) was present on screen at a time. For each trial, participants saw the question "Does the conclusion follow from the premises?" and the words "`Premise:`", "`Premise:`", and "`Conclusion:`" indicating where the statements would appear on the screen. Participants could advance to the next statement by pressing any key. When all three statements had appeared and disappeared, participants were prompted to click buttons reading "yes" or "no" in answer to

[6] For example, two of the variants for the homonym "cabinet" in the Datisi form were as follows:
Variant 1: All cabinets are political appointments. Some cabinets are antique furniture. ∴Some antique furniture are political appointments.
Variant 2: All cabinets are selected by leaders. Some cabinets are cupboards. ∴Some cupboards are selected by leaders.

the argument's validity. They then clicked "Next" to move to the following argument.

Results

Participants judged that the conclusion followed from the premises significantly more often when the middle term was an equivocal polyseme (65%) compared to an equivocal homonym (48%) (unpaired *t*-test: $p < .001$) — see Fig. 3a. This analysis was confirmed by mixed-effects logistic regression fit to trial-level data for the polysemy and homonymy stimuli. The main dependent variable was "validity" response. The core model included independent variables for Ambiguity-Type: Polysemy vs. Homonymy, treatment/dummy coded using Homonymy as the reference level and Syllogism-Form ("Darii", "Datisi", "Dimatis", "Disamis", contrast-coded) along with their interaction. Trial Number (continuous variable, scaled and centered) and RT (continuous variable, scaled and centered) were also included as main effects. The random effect structure included intercepts for participants and the trial list (one of two random orders). We tested for significant of factors in the model using a likelihood ratio test on the χ-squared values from nested model comparisons with the same random effect structure. The best-fitting model was one including main effects and interactions both Ambiguity-Type and Syllogism-Form. This model was a better fit than one that did not include the interaction term ($\chi^2(3) = 18.309, p < .001$) and a better fit than one that did not include Ambiguity-Type ($\chi^2(4) = 45.679, p < .001$). Validity judgments were thus significantly higher for polysemy items than for homonymy items.

Based on the significant interaction between Ambiguity-Type and Syllogism-Form, we split the data into four sets (one for each syllogism) and analyzed them separately. Comparable model comparisons indicated a significant effect of Polysemy (model including a main effect of Ambiguity-Type is a better fit than the nested model which drops Ambiguity-Type) for Darii ($\chi^2(1) = 12.527, p < .001$), Datisi ($\chi^2(1) = 19.287, p < .001$), and Dimatis ($\chi^2(1) = 9.422, p < .003$). However, the model including polysemy is not a better fit than the nested model which drops it for the Disamis stimuli ($\chi^2(1) = 0.702, p = .402$).

We conducted two additional sets of analyses of note. A comparable mixed-effects model comparison confirms that the Homonymy arguments were not judged to be significantly better than the Invalid Fillers — a model including Ambiguity-Type used to predict Homonym vs. Invalid Filler trials was not a better fit than one without the Ambiguity-Type variable ($\chi^2(1) = 0.625, p = .429$).

Finally, we wanted to rule out the possibility that the Uncommon Sense Effect was perhaps driven by a subset of the population who alone endorsed the equivocal polyseme arguments more than the homonym arguments while the remainder of the population did not. We found that this was not the case. While participants vary substantially in their overall rates of argument acceptance — polysemy and homonymy judgments are correlated — the boost to polysemy items is stable across participants (see Fig. 3b).

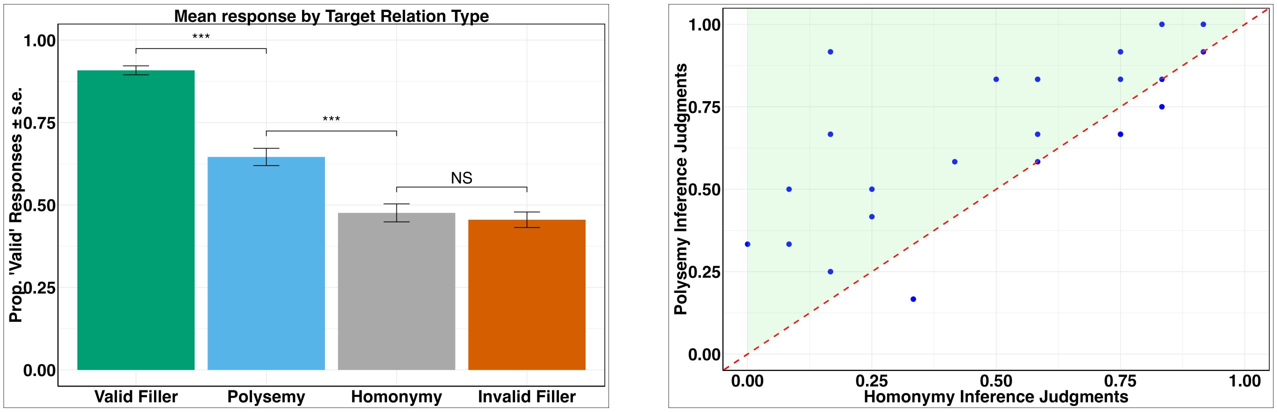
Discussion

These results are consistent with the Underspecification View according to which polysemes and homonyms are represented differently, and that this difference leads reasoners to endorse inferences with equivocal polysemes more often than inferences with equivocal homonyms, the *Uncommon Sense Effect*. This is consistent with an account under which polysemes, but not homonyms, involve underspecified representations shared across senses, and that these underspecified representations can be used in inference. Since inference is a paradigmatically conceptual, non-linguistic process, this evidence also suggests that underspecified, polysemous representations reflect features of non-linguistic cognition that underwrite domain-general conceptual thought.

While the magnitude of the Uncommon Sense Effect is large, and we believe quite general across polysemes, it is worth noting, as mentioned in the Results, that we found a significant effect of polysemy on validity judgments for Darii, Datisi, and Dimatis arguments but not in the case of the Disamis stimuli. Since there were only three unique polyseme middle terms in the Disamis set it is not clear at present whether this is an idiosyncratic by-product of those particular sentences or potentially a real and robust interaction between the Disamis argument form with polysemy inference. More targeted follow ups are needed to this end.

Although we have explained the Uncommon Sense Effect as resulting from a difference in the underlying representations of polysemes and homonyms, our current results leave open two alternative explanations. First, while we controlled for participants' truth judgments of each argument's *conclusion* (Experiment 1), we did not do the same for each argument's *premises*. As a result, the observed effect could in principle be the result of an unintended difference in plausibility across the premises in polyseme/homonym arguments. For example, if reasoners view the premises in arguments containing polysemes more favorably than the premises in arguments containing homonyms, then they may judge the former as more valid than the latter. While we explicitly instructed participants to solely consider the logical validity of each argument and *not* the truth of its premises or conclusion on their own, this possibility can only be conclusively ruled out by a future study which either determines there is no correlation between premise-plausibility and argument-judgment, or explicitly controls for this difference.

Another alternative explanation is that senses' graded semantic relatedness drives the effect. We have interpreted our results to suggest that syllogistic reasoning requires some underspecified representation to re-occur across premises as the middle term. However, it is possible that instead of a common representation, a degree of semantic relatedness across senses is sufficient to draw inferences. For instance, perhaps the mere fact that representations of breakfast foods and breakfast events are highly related (in a way that those of baseball



(a) (b)

Figure 3: Bar graph (panel a) of average validity response split up by word-type, where these types occurred as middle terms (Valid Filler, Polysemy, Homonymy, and Invalid Filler). Judgments are significantly higher (unpaired *t*-tests) for Valid Fillers compared with Polysemy (mean of 91% vs. 65%, $p < .001$), as well as for Polysemy compared with Homonymy (mean of 65% vs. 48%, $p < .001$), but there is no difference in judgments for Homonymy compared with Invalid Fillers (mean of 48% vs. 46%, $p = .564$). Panel b shows the correlation between homonymy and polysemy judgments across participants. Each blue dot represents a single participant, with the green area indicating those demonstrating the Uncommon Sense Effect for polysemy.

bats and animal bats are not) causes reasoners to draw inferences. We plan to address this possibility in a follow-up experiment comparing polysemy arguments to arguments that replace polysemes with distinct words with closely related meanings (e.g., "salt" and "pepper"). The Underspecification View predicts that the Uncommon Sense Effect will not occur for such non-polysemous word pairs, whereas a sense-relatedness explanation predicts that it will.

Another area for future research is the connection between our results and the large body of polysemy research within the parallel-distributed processing (PDP) framework (Armstrong and Plaut (2008, 2016); Rodd, Gaskell, and Marslen-Wilson (2004) — for an overview, see Rodd (2020)). These PDP models' intermediate stage of semantic activation is comparable to the underspecified representations we posit: perhaps they are the same phenomenon described at a different level of explanation or simply modeled using different tools. However, these models have tended to focus on tasks involving isolated words (Rodd, 2020, p. 415), and it is an open question whether they can explain the Uncommon Sense Effect.

Broader Implications

Our results have more far-reaching implications about the nature of inference. On the one hand, and perhaps unsurprisingly, our results show that patterns of deductive inference are not merely sensitive to word *forms*. Otherwise, participants would have drawn inferences over equivocal homonym and polyseme arguments at similar rates. Thus, participants' reasoning does not result from a simple pattern-matching machine which checks that the same string appears in the right places across certain argument forms. On the other hand, and more interestingly, our results also suggest that inferences are not necessarily sensitive to the particular propositional content that is contained in an argument's premises. With some types of words, reasoners seem to endorse inferences that are not truth-preserving. This clashes with a classic version of the computational theory of mind, according to which the mental representations that we use in reasoning are syntactic symbols with semantic content (Pitt, 2022). For if the Underspecification View is correct, the representations we use in reasoning at least sometimes should be distinguished from representations of propositional content.

Our finding also bears on a core question in cognitive science: the nature of concepts. There is disagreement among those who endorse the Underspecification View about whether underspecified representations are restricted to language processing or can be used in conceptual thought. For example, Pietroski (2018, 2025) claim that the underspecified representation of a polyseme is confined to the linguistic system and that mapping a polysemous word to a concept requires accessing a determinate sense (e.g., BREAKFAST_{food}). In contrast, Quilty-Dunn (2021) argues that a polyseme's underspecified representation occurs at the level of conceptual thought (e.g., $\text{BREAKFAST}_{underspecified}$ — see also Chomsky (1995)). If concepts are elements of inference as frequently supposed in the literature (Carey, 2009; G. Murphy, 2002; Peacocke, 1992), then, given that our results show that humans can use underspecified representations in inference, it follows that underspecified representations are conceptual rather than merely linguistic. We can *think* without thinking about anything in particular. This result complicates views of concepts according to which they are both the bearers of content (and can thus, at least sometimes, refer to objects in the world) and the mental particulars used in inference. These two central desiderata may need to be separated.

Acknowledgments

The authors thank the CUNY StarLab. This work was supported by PSC-CUNY Award 67607-00 55.

References

Armstrong, B. C., & Plaut, D. C. (2008). Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 30).

Armstrong, B. C., & Plaut, D. C. (2016). Disparate semantic ambiguity effects from semantic processing dynamics rather than qualitative task differences. *Language, Cognition and Neuroscience*, *31*(7), 940–966.

Beretta, A., Fiorentino, R., & Poeppel, D. (2005). The effects of homonymy and polysemy on lexical access: An meg study. *Cognitive Brain Research*, *24*(1), 57–65.

Carey, S. (2009). *The origin of concepts*. Oxford University Press. doi: 10.1093/acprof:oso/9780195367638.001.0001

Chomsky, N. (1995). Language and nature. *Mind*, *104*(413), 1–61.

Devitt, M. (2021). Semantic polysemy and psycholinguistics. *Mind & Language*, *36*(1), 134–157.

Evans, J. S. B., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & cognition*, *11*(3), 295–306.

Falkum, I. L., & Vicente, A. (2015). Polysemy: Current perspectives and approaches. *Lingua*.

Fodor, J., & Lepore, E. (2002). *The compositionality papers*. Oxford University Press.

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of memory and language*, *29*(2), 181–200.

Frisson, S. (2009). Semantic underspecification in language processing. *Language and linguistics compass*, *3*(1), 111–127.

Frisson, S. (2015). About bound and scary books: The processing of book polysemies. *Lingua*, *157*, 17–35.

Frisson, S., & Pickering, M. J. (1999). The processing of metonymy: evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(6), 1366.

Frisson, S., & Pickering, M. J. (2007). The processing of familiar and novel senses of a word: Why reading dickens is easy but reading needham can be hard. *Language and cognitive processes*, *22*(4), 595–613.

Hasson, U., & Glucksberg, S. (2006). Does understanding negation entail affirmation?: An examination of negated metaphors. *Journal of Pragmatics*, *38*(7), 1015–1032.

Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language*, *45*(2), 259–282.

Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, *20*(1), 1–24.

Klepousniotou, E., Titone, D., & Romero, C. (2008). Making sense of word senses: the comprehension of polysemy depends on sense overlap. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1534.

Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge university press.

McElree, B., Frisson, S., & Pickering, M. J. (2006). Deferred interpretations: Why starting dickens is taxing but reading dickens isn't. *Cognitive Science*, *30*(1), 181–192.

Murphy, E. (2021). *Linguistic representation and processing of copredication*. Unpublished doctoral dissertation, UCL (University College London).

Murphy, G. (2002). *The big book of concepts*. MIT press.

Nida, E. A. (1979). *A componential analysis of meaning: An introduction to semantic structures*. Walter de Gruyter GmbH & Co KG.

Peacocke, C. (1992). *A study of concepts*. MIT Press.

Pietroski, P. M. (2018). *Conjoining meanings: Semantics without truth values*. Oxford University Press.

Pietroski, P. M. (2025, 01). One word, many concepts: Endorsing polysemous meanings. In *The oxford handbook of contemporary philosophy of language.* Oxford University Press. Retrieved from `https://doi.org/10.1093/oxfordhb/9780192856852.013.34` doi: 10.1093/oxfordhb/9780192856852.013.34

Pitt, D. (2022). Mental Representation. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford encyclopedia of philosophy* (Fall 2022 ed.). Metaphysics Research Lab, Stanford University. `https://plato.stanford.edu/archives/fall2022/entries/mental-representation/`.

Quilty-Dunn, J. (2021). Polysemy and thought: Toward a generative theory of concepts. *Mind & Language*, *36*(1), 158–185.

Reverberi, C., Pischedda, D., Burigo, M., & Cherubini, P. (2012). Deduction without awareness. *Acta psychologica*, *139*(1), 244–253.

Rodd, J. M. (2020). Settling into semantic space: An ambiguity-focused account of word-meaning access. *Perspectives on Psychological Science*, *15*(2), 411–427.

Rodd, J. M., Berriman, R., Landau, M., Lee, T., Ho, C., Gaskell, M. G., & Davis, M. H. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory & Cognition*, *40*, 1095–1108.

Rodd, J. M., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of memory and language*, *46*(2), 245–266.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive science*, *28*(1), 89–104.

Srinivasan, M., Berner, C., & Rabagliati, H. (2019). Children use polysemy to structure new word meanings. *Journal of Experimental Psychology: General*, *148*(5), 926.

Srinivasan, M., & Rabagliati, H. (2015). How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua*, *157*, 124–152.

Srinivasan, M., & Snedeker, J. (2011). Judging a book by its cover and its contents: The representation of polysemous and homophonous meanings in four-year-old children. *Cognitive psychology*, *62*(4), 245–272.

Zehr, J., & Schwarz, F. (2018). *Penncontroller for internet based experiments (ibex).* Online. Retrieved from https://doi.org/10.17605/OSF.IO/MD832