Journal of Memory and Language 91 (2016) 117–140

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Shared and distinct mechanisms in deriving linguistic enrichment ☆

Lewis Bott [a,*], Emmanuel Chemla [b]

[a] School of Psychology, Cardiff University, United Kingdom
[b] Laboratoire de Sciences Cognitives et Psycholinguistique (ENS, EHESS, CNRS), Département d'Etudes Cognitives, Ecole Normale Supérieure, PSL Research University, France

ARTICLE INFO

Article history:
Received 15 September 2015
revision received 11 April 2016
Available online 21 May 2016

Keywords:
Pragmatics
Implicatures
Structural priming

ABSTRACT

Meanings of basic expressions can be enriched by considering what the speaker could have said, but chose not to, that is, the *alternatives*. We report three priming experiments that test whether there are shared enrichment mechanisms across a diverse range of linguistic categories. We find that quantifier, number, and *ad hoc* enrichments exhibit robust priming within their categories and between each other. Plural enrichments, in contrast, demonstrate within-category priming but no between-category priming. Our results demonstrate that (1) enrichment typically thought of as pragmatic or semantic can be primed in the same way as syntactic structures and (2) there are mechanisms that are shared across different enrichment categories, and that some phenomena (e.g., plurals) are excluded from this class. We discuss the implications of our findings for psychological models of enrichment, theories of individual categories of enrichment, and structural priming.

Introduction

Understanding a sentence requires mapping the words that were uttered to appropriate meanings and then piecing them together according to a grammar. But it also requires considering words the speaker did not utter, but could have. The listener must consider the alternatives to what the speaker said and incorporate them into the message expressed by the spoken words. Consider the examples below.

1. A: Have you met Dave, Jane's new boyfriend? He's intelligent and handsome.
B: Well, he's intelligent.
⇒ *he's not handsome*
2. There's a pen on the table.
⇒ There's a pen *and nothing else* on the table
3. Some of the children are in the classroom
⇒ *Not all* of the children are in the classroom
4. I've got two children.
⇒ I've got two children *but no more than two*.
5. John's essay was acceptable.
⇒ John's essay was acceptable *but not excellent*.

In (1), the meaning of B's words is that Dave is clever. However, in the right context, B's utterance communicates more than this. Speaker A would be licensed to infer that B believes Dave is *not handsome*. In order to derive the additional meaning, Speaker A might reason as follows.

☆ This research received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n.313610 and was supported by ANR-10-IDEX-0001-02 PSL* and ANR-10-LABX-0087 IEC.

* Corresponding author at: School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 3AT, United Kingdom.
E-mail address: BottLA@cardiff.ac.uk (L. Bott).

http://dx.doi.org/10.1016/j.jml.2016.04.004
0749-596X/

First, she would consider what B could have said (the alternatives), such as, "Yes, you're right." or "Yes, he's intelligent and handsome". Then, she could reason that since B did not say those alternatives, and B was in a position to make that judgment (e.g., B had met Dave), B must not believe the alternatives. Put simply, if B had meant any of these alternatives she could have said them, and since she did not, it can be inferred that she does not believe them. The enrichment shown in the other examples can all be derived using similar reasoning: In (2), the speaker could have described the other objects on the table but because s/he did not, the listener can infer that there are no other objects on the table; in (3), the speaker could have used other, more informative quantifiers instead of *some*, such as *all*, so *not all* can be derived; in (4), the alternatives to *two* are *three, four, five,* etc. and so these can be negated to give *no more than two*; and in (5), John's essay could have been *good* or *excellent* but since the speaker did not say that it was, the listener can infer that it is neither *good* nor *excellent*.

The classic reasoning above is inspired by Grice (1975), who viewed such enrichments as a natural consequence of speakers and listeners cooperating in dialog. Recent debates have tried to evaluate the respective role of grammatical and domain general reasoning processes to derive such inferences (see, Chierchia, Fox, and Spector (2012), for a strikingly unorthodox position). But as a starting point, most current proposals agree that they form a natural class of phenomenon, known as scalar implicatures, involving the competition with alternatives. In this paper we question the extent to which different categories of enrichment should be grouped together on mechanistic grounds. We test whether there are shared reasoning processes that apply across enrichments, as in the reasoning sketched above, or whether each enrichment, or category of enrichments, uses a set of specialized procedures.

Our approach was to test whether enrichments can be *primed* across expressions. If different sorts of enrichments can prime each other, there must be an abstract mechanism that is shared between them. By testing which enrichments prime each other and which don't, we can specify what the common mechanism might be.

Categories of enrichment

We are concerned with enrichments that arise via the use of alternatives; hence we refer to the general phenomena as *enrichment-via-alternatives* (EVAs).[1] The EVAs shown in Examples (1)–(5) involve a variety of different linguistic forms. They were chosen to illustrate how EVAs function in general but also because they are representative of three categories of EVAs around which there is debate about whether a common mechanism is used for their derivation. These categories form the basis of our hypothesis and our experimental materials. Here we introduce the different categories.

Quantifiers

Sentences involving quantifiers such as *some* are generally taken to have at least two interpretations: a weak reading, and a strong reading. The weak reading is typically consistent with an *at least* meaning, as in (6) below:

(6) If some of the children are in classroom, I'll shut the window.

Here, it is clear that if at least one, and possibly all, of the children are in the classroom, the window will be closed. This can be compared with the strong meaning, illustrated in (3), which conveys an *only*-like reading. In (3) the speaker likely means that at least one child, but not all of them, are in the classroom. The argument from Horn (1972), Gazdar (1979), Levinson (1983) and many others since is that the weak reading can be enriched by combining its basic meaning with the negation of some alternative. In the *some* case, the weak reading, *at least one,* is combined with the negation of the alternative constructed by replacing *some* with *all*, to therefore generate the *at least one, but not all,* reading.

Quantifier enrichments are seen as prototypical examples of EVAs in the theoretical (e.g., Horn, 1972; Levinson, 1983) and the experimental literature (e.g., Bott & Noveck, 2004; Breheny, Katsos, & Williams, 2006; Degen & Tanenhaus, 2015; Huang & Snedeker, 2009; Grodner, Klein, Carbary, & Tanenhaus, 2010). We included quantifiers as a distinct category of EVAs in our experiments to act as a kind of benchmark against which other (potential) EVAs could be compared.

Numerals

As with quantifiers, sentences with number words can also be described as having a weak and a strong meaning. The strong meaning corresponds to an exact sense, as in (4), and corresponds to the most prominent meaning of cardinal terms. The weak (or at least) meaning shows up in examples such as (7); here the speaker uses "two" to mean two or more (at least two), since having three children would not prevent the applicant from receiving support.

(7) Parents with two children will be eligible for financial support.

The relationship between the weak and strong number meanings can be seen in the same way as that between the weak and strong quantifiers (e.g., Horn, 1972, 1989). Just like quantifiers, the numbers form an entailment scale, in which numbers at the upper end of the scale entail those at the lower end (that is, whenever sentences with numbers at the upper end of the scale are true, so are sentences with numbers at the lower end). If the numbers can be thought of as having a lexical meaning that corresponds to *at least N*, that is, a weak meaning, then a strong mean-

[1] We use the term *EVA* rather than *implicature* because we prefer to remain theory-neutral about how EVAs arise. While EVAs are often described using a form of Gricean reasoning, there is a lively debate as to whether this reasoning really is a by-product of rational conversations or has to be understood as part of the language decoding system (see, e.g., Chierchia et al., 2012). We need not take part in this debate, once we recognize that all current accounts rely on the same two key pieces: a set of alternatives and a computational system that deal with these alternatives.

ing can be derived by combining the weak meaning with the negation of its alternatives, that is, $N+1$, $N+2$, $N+3$, etc. For example, the strong meaning of "two" can be formed by combining its weak meaning (the lexical meaning), *at least two*, with *not at least three (=less than 3)*, *not at least four (=less than 4)*, etc. to give *at least two but less than 3*, i.e. the *exact* meaning.

Numbers and quantifiers are thus plausibly derived in the same way. They also share a number of distributional properties, such as being suspended under negation (Gazdar, 1979; Horn, 1989; Chierchia, 2004). But they also exhibit divergent behavior. For example, consider (8) and (9) (from Huang, Spelke, & Snedeker, 2013) below

(8) Everyone who ate some of their berries felt fine.
(9) Everyone who ate two of their berries felt fine.

In (8), *some* typically receives a weak reading (eating at least some of the berries) whereas as in (9), *two* typically receives a strong reading (eating exactly two is fine, but eating more than that probably isn't). This has led some researchers to suggest accounts of number meanings in which the strong meaning (the exact meaning) is the lexically determined meaning, and the weak meaning is derived through further computations (e.g., Breheny, 2008; Carston, 1998; Geurts, 2006; Horn, 1992; Sauerland, 2003), quite the opposite of the EVA account suggested above in which the weak meaning is stored lexically, and the strong meaning is derived through enrichment. This work suggests that numbers are not like quantifier EVAs, possibly not EVAs at all, and that they involve at least some distinct psychological processes (for a more extensive review, see Kennedy (2013) and Spector (2013)).

Ad hoc. Many expressions can be enriched using alternatives that are entirely contextually determined, so that the consequent enrichment is made on an *ad hoc* basis. For example, the *not handsome* enrichment that arises in (1) is entirely dependent on A's previous statement about Jane's new boyfriend, Dave, being intelligent and handsome. Had A described Dave's qualities as being intelligent and rich, a different enrichment would have arisen from B's statement; namely that Dave was *not rich*.

Ad hoc enrichments share many of the same properties as the quantifier and the numeral enrichments (Hirschberg, 1991). Most importantly, the same type of ambiguity arises between weak and strong meanings. In particular, in the absence of a context, "he's intelligent" may simply convey a weak meaning, something like *at least intelligent*, which includes neither the *not handsome* nor the *not rich* enrichment.

Ad hoc EVAs look very similar to that of the numbers and the quantifiers but there could also be a dichotomy between the processing of quantifiers/numbers on the one hand and *ad hoc* EVAs on the other, since there are important differences between the two – the nature of the alternatives in particular is necessarily context dependent for *ad hoc* EVAs, while it is less so for quantifiers and numbers (*4* and *all* are privileged alternatives for *3* and *some*, in a way that *at most 3* and *not all* can never be, regardless of the context).

Shared and distinct enrichment mechanisms

The foregoing discussion introduced three commonly studied categories of EVAs. The EVAs are similar in that a trigger expression can give rise to two distinct but related interpretations: the strong interpretation and the weak interpretation. Moreover, in each case, the strong interpretation is plausibly derived from the weak interpretation using the same set of mechanisms. Although there are many explanations for how the enrichment arises, most involve something like the following, which we refer to as the "core account": (i) the listener computes the weak meaning of the phrase, (ii) recognizes that an alternative phrase could have been used, but that it wasn't, (iii) negates the alternative and combines it with the weak meaning (assuming the speaker is judged to be knowledgeable). Modern and developed theories can be found in van Rooij and Schulz (2004), Sauerland (2004), Franke (2011), Chierchia et al., (2012), a.o.

Our study tests whether the mechanisms described by the core account are indeed shared across EVAs, and to what extent. At one extreme, the same mechanisms could be used to derive quantifier, number and *ad hoc* enrichments. While there is variability across EVAs in the rate of enrichments (e.g., van Tiel, van Miltenburg, Zevakhina, & Geurts, 2014), this could be explained by differences in the frequencies of the alternatives, or some other factor linked to the idiosyncratic properties of the trigger expression. At the other extreme, each EVA could have its own distinct set of mechanisms. For example, *ad hoc* EVAs might be derived using procedures very similar to the core account; numbers might have a lexical entry corresponding to the strong (exact) interpretation, which is then modified to derive the weak interpretation; and the quantifiers might have both strong and weak meanings lexicalized, with the appropriate meaning on any given occasion determined by frequency and probabilistic factors, much like standard polysemy (see Tomlinson, Bailey, and Bott (2013) for a suggestion along these lines). Another possibility is that there are multiple versions of the mechanisms described by the standard account, one for each EVA: mechanisms specialized in retrieving and negating *ad hoc* alternatives; mechanisms specialized in retrieving and negating number alternatives; and mechanisms specialized in retrieving and negating quantifier alternatives. Finally, between the extremes, some of the mechanisms might be shared and other distinct. For example, all three EVAs could use the same process for negating the alternatives and combining the result with the basic meaning, but the retrieval of the alternatives may be different: for *ad hoc* EVAs it *necessarily* involves context specific processes while the (standard) alternatives for quantifiers and numbers may not.

Existing data is not able to distinguish between these possibilities. Despite the similarities in the distributions of numbers, quantifiers and *ad hoc* EVAs, which point to a shared set of mechanisms, there are also dissimilarities, such as that between the numbers and the quantifiers (e.g., Breheny, 2008), which point to distinct mechanisms.

More fundamentally, it is not clear whether this kind of evidence can categorically answer questions about shared psychological mechanisms. The basic problem is that there are multiple psychological mechanisms that can give rise to similar kinds of distributional (or psychological) behavior. As an example consider the *exact vs at least* accounts of numbers considered above. Here, there are different representational processes hypothesized to account for very similar distributions. Indeed, even if the distributions were identical it would still be possible for distinct psychological mechanisms to underpin quantifier and number enrichment. Similarly, one may observe differences between EVAs but this cannot exclude that some mechanisms be shared. For example, the developmental literature now converges towards saying that variability in how children interpret different EVAs concern their knowledge of alternatives, while the rest of the system could be unaltered (Barner, Brooks, & Bale, 2011; Tieu, Romoli, Zhou, & Crain, in press). The issue is that the data has been correlational - correlations between the distribution of different EVAs and correlations between processing patterns - whereas to establish whether there are shared mechanisms requires establishing whether enrichment of one expression can cause the enrichment of another. In the reminder of the article we present three experiments that test this. Our goal is primarily to identify whether there are any levels of shared representation that mediate inferencing across the EVAs considered above.

Priming enrichment

The current study used a structural priming paradigm (e.g., Raffray & Pickering, 2010). Structural priming occurs when participants adopt a particular linguistic structure on one trial (the *prime*) and then adopt the same structure on a subsequent trial (the *target*). For example, in Bock (1986), participants repeated a prime sentence that could be in active form (e.g., "One of the fans punched the referee") or passive form (e.g., "The referee was punched by one of the fans"), and then had to describe a picture. Participants were more likely to describe a picture in passive form after they had repeated a passive prime sentence than an active prime sentence. Similar effects have been shown in a huge range of production tasks, including written production (e.g., Pickering & Branigan, 1998) and dialog (Branigan, Pickering, & Cleland, 2000), as well as single participant description tasks (e.g., Raffray, Pickering, Cai, & Branigan, 2014). Priming can also facilitate comprehension in sentences (Thothathiri & Snedeker, 2008), and influence the final analysis of globally ambiguous sentences, such as scopally ambiguous sentences (Raffray & Pickering). Our task uses a sentence–picture matching task, modeled on Raffray and Pickering, in which participants are constrained to derive either a strong or a weak interpretation of an expression on a prime trial, and then make a judgment about whether to enrich the expression on a subsequent target trial. Priming of enrichment would be shown by a greater proportion of strong interpretations of the target sentence after a strong prime than after a weak prime.

The results of priming experiments have been used to argue for the existence of representations that are abstracted away from superficial properties of the sentence (e.g., Bock, 1986; Pickering & Ferreira, 2008). The basic premise of structural priming is that people have language representations that are constructed from part-of-speech forms, such as nouns, verbs, and prepositions and constituents organized from those forms, such as noun phrases, verb phrases, propositional phrases, and that producing or comprehending a sentence activates particular constructions. These constructions then remain active across trials so that the next time a suitable sentence is encountered, the primed construction has an advantage over other potential structures, and the sentence is produced or comprehended according to the primed structure.

The logic of our approach is similar to that described above. If a strong prime interpretation of one EVA sentence causes a greater number of strong target interpretations of a different category, this would be evidence for EVA mechanisms that operate beyond sentence specifics, and indeed beyond EVA categories. For example, the core account assumes a mechanism that uses the negation of the alternatives, as in *not handsome* in (1) or *not all* in (2). If it is the same mechanism that negates the alternatives across different categories of EVA, it should be possible to prime the mechanism, such that the probability of negating alternatives (and so deriving the enrichment) is greater after a strong prime trial than after a weak prime trial.

Experimental overview

In all of our experiments participants saw a sentence and had to match the sentence with one of two pictures. The sentences referred to the presence of symbols in a set, such as "All of the shapes are diamonds." In the experimental trials, the sentences invited enrichment. However, because the enrichment was optional, participants could interpret the sentence in its basic or enriched form. This meant that the sentences could have either a *weak* meaning (without enrichment) or a *strong* meaning (with enrichment). For a given sentence, three types of pictures were possible: (a) *false pictures*, that made both readings false, (b) *weak pictures*, that made the weak reading true but the strong reading false, and (c) *strong pictures* that made both readings true. Pictures were arranged in various combinations to form *prime* trials and *target* trials. There were two types of prime trials. First, *weak primes*, which displayed a false picture and a weak picture, so that participants would click on the weak picture and access the weak reading. Second, *strong primes*, which displayed a weak picture and a strong picture. We reasoned that participants would access the strong reading (the one that makes the two pictures different in a relevant way) and click on the strong picture. The prime trials were designed to force a particular interpretation of the sentence. There were strong and weak trials for each of the EVA categories used in the experiment. Figs. 1 and 3 show example prime sentence–picture pairings.

In the target trials, participants read another experimental sentence and saw two more pictures. One of the pictures was a weak picture, and the other picture was a box with "Better Picture?" written inside it. Participants were instructed that the "Better Picture" option should

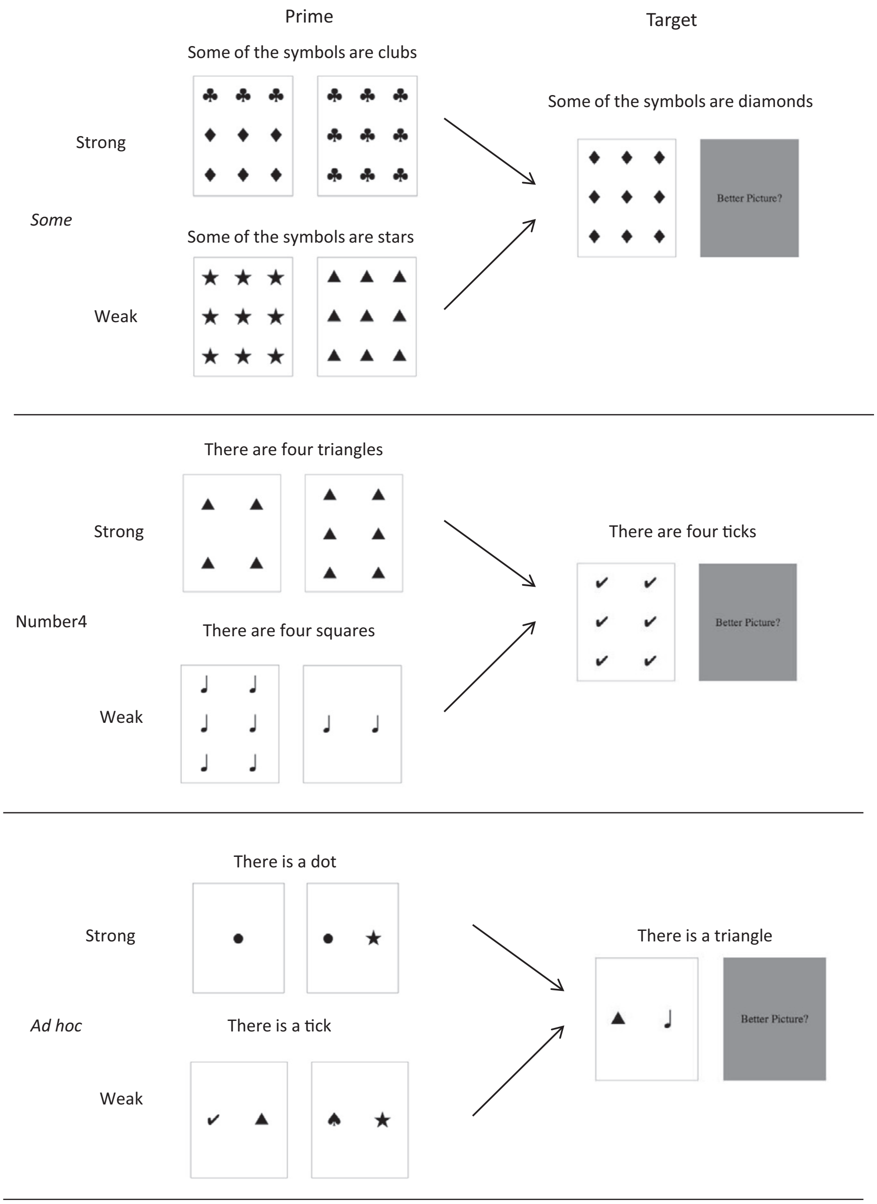

Fig. 1. *Example stimuli for Experiment 1.* Participants see a prime followed by a target. The prime (left column) consists of a sentence and two pictures, and the target (right column) one picture and the "Better Picture?" option. Each panel illustrates stimuli from one of the three EVA categories: *some*, number4, and *ad hoc*. Strong trials (upper half of the panel) and weak trials (lower half of the panel) are shown for each EVA category. Within-category trials consist of prime and target from the same category whereas between-category trials consist of prime and target from a different category.

be selected if they did not feel that the other picture sufficiently captured the sentence meaning (we modeled the "better picture" method on Huang et al. (2013)). Figs. 1 and 3 show examples of the target trials. We expected that participants should click on the weak picture if they accessed the weak reading, and opt for the "Better Picture" option if they accessed the strong reading. Target trials immediately followed prime trials. Consequently, priming of the enriched meaning would be observed when a participant selected the strong interpretation option more often after the strong prime than after the weak prime (and vice versa).

Each of the experiments used sentences from multiple EVA categories. For example, Experiment 1 used *some* sentences and number sentences. There were prime and target sentences for each category. There were therefore two different forms of priming, within-category priming, in which prime and target were of the same type, such as a *some* prime preceding a *some* target, and between-category priming, in which prime and target were of different types, such as a *some* prime and a *number* target.

Evidence of shared mechanisms across EVAs would be shown by significant between-category priming. Evidence of within-category priming is also of interest, however,

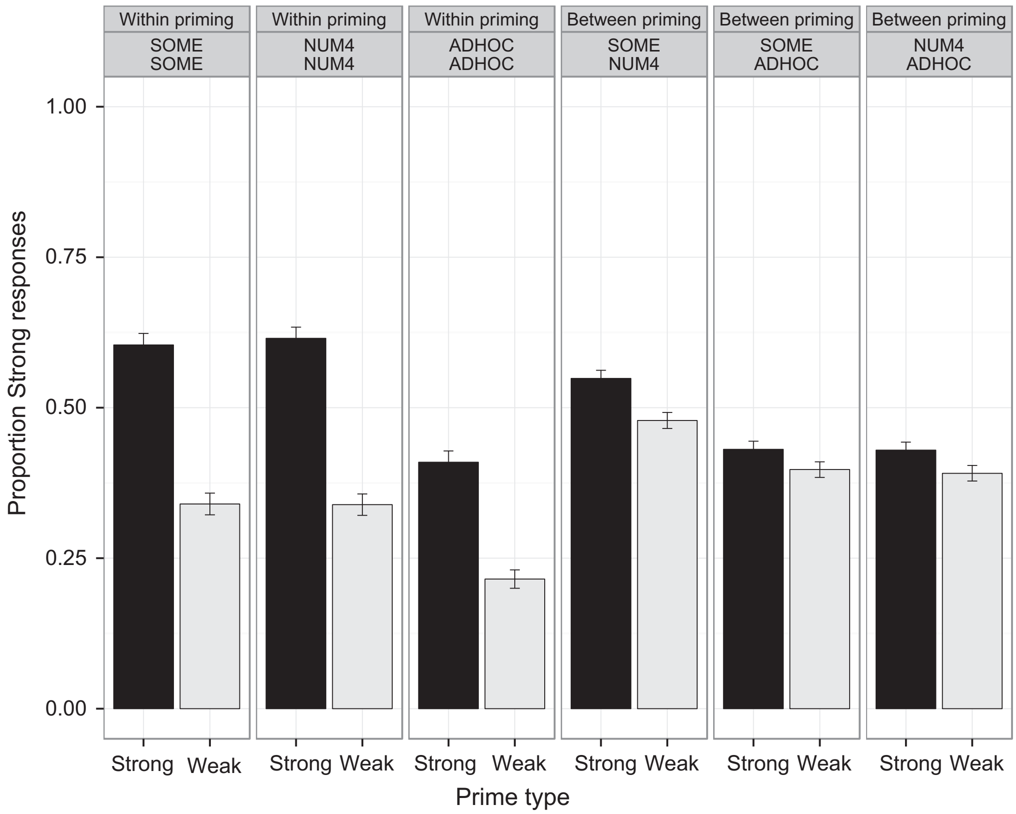

Fig. 2. *Experiment 1 results.* Priming is shown by the difference between the strong and weak bars for each panel. Within-category priming is shown in the left three panels, and between-expression priming in the right three panels. Note that the between-expression priming groups are pooled across priming direction (e.g., "SOME NUM4" is the combination of *some* → number4 and number4 → *some* trials).

for the following reasons. First, within-category priming would demonstrate that at least some enrichment mechanisms were primeable, that is, remain active across time and linguistic material. We would consequently have stronger grounds for arguing that there were no shared enrichment mechanisms if we were not to observe between-category priming. Second, a comparison of the within-category effect to the between-category effect could provide information about which mechanisms were being primed. For example, if only shared mechanisms were primed, then there should be no difference between within and between-category priming effects.

Experiment 1

In Experiment 1 we tested *some*, number and *ad hoc* EVAs. Examples are shown in Fig. 1. Within-category trials involved a prime and target from the same category, that is, *some* → *some*, number4 → number4, and *ad hoc* → *ad hoc*. Between-category trials involved a prime and target from different categories, such as *some* → number4, or *ad hoc* → *some*. If enrichment can be primed at all, we would expect within-category priming. If the numbers, *some* and *ad hoc* EVAs share enrichment mechanisms we would expect them to prime each other, so that a strong *some* prime, for example, leads to a greater proportion of strong number responses.

Method

Participants

Two hundred participants were recruited using Amazon Turk. Of these, 13 were removed because they did not declare English as their native language. The data from the remaining 187 were used in the experiment.

Materials

Each trial involved a sentence presented above two pictures. Participants had to match the sentence to one of the pictures. For experimental trials, the sentence was constructed using one of three frames: (i) Some of the symbols are [symbol] (ii) There are four [symbol] (iii) There is a [symbol].[2] The symbols were one of diamonds, clubs, ticks, spades, hearts, squares, stars, circles, notes, or triangles.

[2] The *some* and the number/*ad hoc* sentence frames differed in that the EVA trigger was in subject position for *some* sentences whereas it was in object position for numbers sentences. One likely effect of this would be to elevate the rate of enrichment for *some* sentences and to suppress them for number sentences (see e.g., Breheny et al., 2006). This is useful in our case because the two types of scale probably have different enrichment rates (*some* having a low enrichment rate and the numbers a high enrichment rate) and the difference in sentence structure brought the two enrichment rates to the center of the scale. A further effect might be to make the sentences relatively dissimilar, thereby reducing between-category priming. However, since we did find priming (as the upcoming results demonstrate), the difference in sentence structure makes our result even stronger.

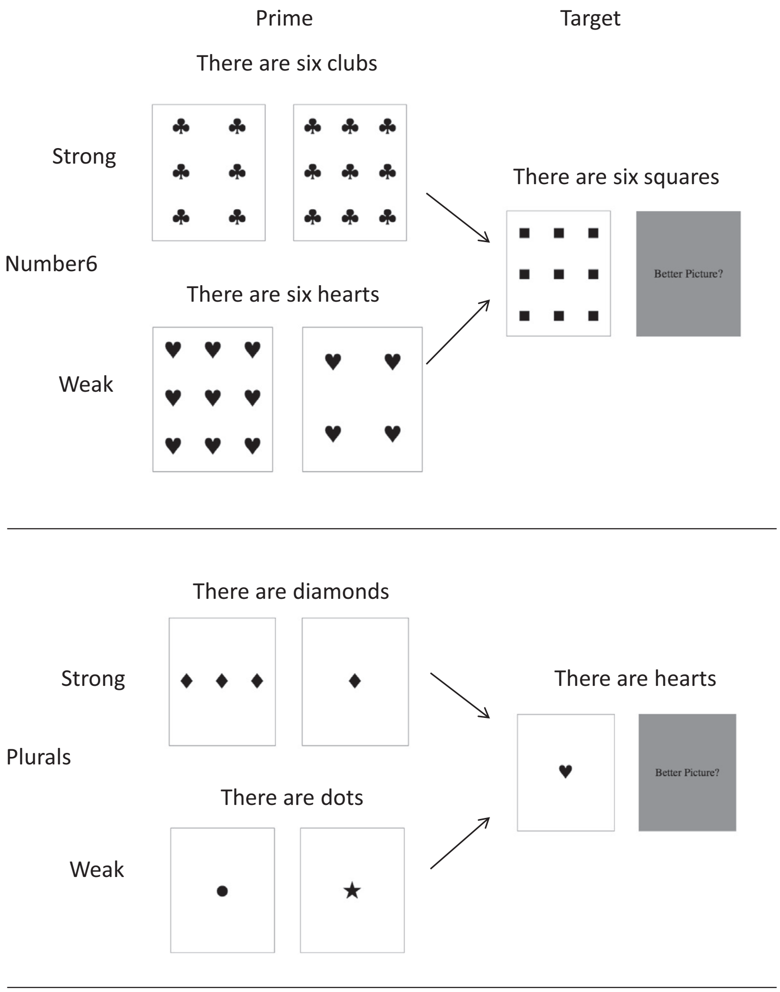

Fig. 3. *Example stimuli from Experiments 2 and 3.* The upper panel shows stimuli from number6 EVAs and the lower panel from plurals.

Pictures consisted of rectangles containing either symbols or the text "Better Picture?" For prime trials, both pictures contained symbols. For target trials, one picture contained symbols and the other, "Better Picture?"

Pictures with symbols could be strong, weak or false. Weak prime trials involved a weak and a false picture, and strong prime trials involved a strong and a weak picture.

For *some* trials, strong pictures involved three symbols of one type and six of another type. The predicate in the sentence always matched the minority symbol type. For example, if the sentence was, "Some of the symbols are diamonds," the strong picture involved three diamonds and six of another symbol, e.g., spades. Weak pictures involved nine symbols of the type that matched the predicate. False pictures involved nine symbols of a type that did not match the predicate. For number4 trials, strong pictures involved four symbols that matched the predicate, and weak pictures involved six symbols that matched the predicate. False pictures involved two symbols that matched the predicate. Finally, for *ad hoc* trials, weak pictures contained two symbols, one of which was consistent with the predicate, and strong pictures contained a single symbol that was consistent with the predicate. False *ad hoc* pictures contained two different symbols, neither of which matched the predicate.

We also included filler trials linked to each prime-target combination. These involved sentences that were more informative than the basic expressions used in the experimental trials (given the relevant weak picture). There were *all* sentences (an alternative to *some*), e.g., "All the symbols are diamonds," which was more informative given the weak *some* picture; *six* sentences (an alternative to *four*), e.g., "There are six symbols;" and *double* sentences (an alternative to *ad hoc*), e.g., "There is a diamond and a square." Each occurred in three forms (1) a weak picture with symbols that did not match the predicate in the sentence, and a "Better Picture?" option (2) a weak picture with symbols that matched the predicate, and a "Better Picture?" option, and (3) a weak picture with symbols that matched the predicate, and a strong picture. These items served to highlight that alternative to the participant, thereby elevating the overall rate of enrichment, and to prevent the participant adopting response strategies without fully processing the sentence.

In addition to the *ad hoc* filler trials we included *ad hoc bias* trials at the start of the experiment. The bias trials were designed to elevate the overall proportion of enrichment responses (in pilot studies without the bias trials we found that participants responded at a very low rate of enrichment for the *ad hoc* trials). Bias trials involved the standard *ad hoc* sentence, "There is a [symbol]," and occurred in two forms. In one form, the sentence was paired with a symbol picture and a "Better Picture?" option, and in the other, there were two symbol pictures. In the "Better Picture?" trials, one picture contained a symbol that did not match the predicate and the other was the "Better Picture?" option. In the symbol picture trials, one picture was the strong *ad hoc* picture (a single symbol that matched the predicate) and the other was a picture with a single symbol that did not match the predicate. The idea behind the bias trials was to facilitate participants in imagining what the appropriate "better picture" might be for the enriched expression.

Design

There were three types of enrichment category (*some,* number4, *ad hoc*). For each, there were two prime types, a strong prime trial and a weak prime trial, and a target trial. There were consequently 3 (enrichment category) × 2 (prime type) = 6 distinct prime trials and 3 (enrichment category) target trials. We fully crossed these to form 18 distinct prime-target combinations. Of these, 6 involved primes and target from the same enrichment category (within-expression trials), such as, *some* prime → *some* target, and 12 involved primes and target from different categories (between-expression trials), such as *some* prime → number4 target.

There were two primes for every target, so that the experimental units were triplets of trials, such as, strong *some* prime, strong *some* prime, *some* target. This was done to boost the effect of the prime. There were 4 examples of each prime-target combination. Consequently there were 4 (examples) × 18 (prime-target combinations) × 3 (triplets) = 216 experimental trials.

There were a further 36 filler trials, 12 per enrichment category, with equal numbers of the three filler types described above, and 16 *ad hoc* bias trials.

Randomisation and counterbalancing

All participants saw the same set of trials. The symbol in the sentence and target image or in the false image was picked at random, with replacement across trials. These trials, in a triplet of prime-prime-target or individual fillers where then administered in a random order to each participant.

For prime trials there was a correct response option. For weak primes this was the weak picture. For strong primes this was the strong picture but for an indirect reason: in the presence of both a weak picture and a strong picture, participants could not make a non-arbitrary choice solely based on the truth conditions of the weak interpretation which is true in both cases, hence the strong reading is a favored option in that it provides a non-arbitrary way to resolve the task. The position of the correct response was counterbalanced across trials so that for half the trials it was on the right and for half it was on the left. Furthermore, for half the trials the correct response was the same side as the previous trial and for half the trials it was on the opposite side. For target trials there was no correct answer but the "Better Picture?" option was always on the right.

Procedure

Participants were instructed to click on the picture that "best matched the sentence." They were given two simple examples, one involving *many*, as in "Many of the symbols are stars," and one involving *above*, as in "There is a diamond above a square." The latter involved the "Better Picture?" option. They were instructed to select this option if they thought there was "a picture that better matched the sentence."

Responses were selected by clicking with the mouse on a button beneath the pictures.

Results

Data treatment

Each target trial was preceded by two prime trials. Without participants correctly responding to the prime trials we could not be sure that they had derived the correct interpretation of the prime sentence. We therefore removed all target responses that were not preceded by the two correct prime responses (in common with Raffray and Pickering (2010), and many others). In this experiment, 875 out of 13,360 target responses were removed because of incorrect prime responses. Of the 875, 279 were *ad hoc* targets, 273 were number targets and 323 *some* targets.

Analysis procedure

We analyzed our data by modeling response-type likelihood using logit mixed-effect models (Jaeger, 2008). The random effects structure included random intercepts and slopes for all repeated measures factors (we had no between-subject factors). Analyses were conducted using the lme4 (Bates, Maechler, & Bolker, 2011) and languageR libraries (Baayen, 2011) for the R statistics program (R Core Team, 2014). β values, standard errors, Z-values, and p-values are shown in the tables accompanying the experiment together with R pseudo-code describing the models. Treatment and sum coding were used as appropriate and the reference levels are stated in the text. The Appendix shows raw means for each cell in the design.

In all of our experiments we start with an analysis involving all of the data, in which we assessed within-expression priming, between-expression priming and the interaction between the two. To gain a more detailed picture we then restricted the analysis to within-expression trials only and between-expression trials only. The dependent measure was the log odds of choosing a strong over a weak response on target trials.

Analysis

Fig. 2 shows the data for Experiment 1. The figure is divided into six panels. The first three show responses to targets when the prime and target were of the same category, i.e., within-category priming trials. There is one panel for *some*, one for number4 and one for *ad hoc* targets respectively. The separate bars within each panel refer to the value of the prime, either strong or weak. The large difference between the strong and the weak primes suggest a substantial within-category priming effect. The second three panels show responses to targets where the target and prime are of a different type. Here, each panel refers to one of three between-category prime-target combinations, *some* ↔ number4, *some* ↔ *ad hoc*, or *number4* ↔ *ad hoc*. For these panels, we combined targets from the two sorts of relevant between-category trials, i.e., we ignored the direction of the prime-target combination (the raw mean for each cell of the design is shown in the Appendix). Thus the *some* ↔ number4 panel consists of responses to number4 targets from *some* (prime) → number4 (target) trials combined with *some* targets from *number4* (prime) → *some* (target) trials. Similarly, the *some* ↔ *ad hoc* panel consists of *ad hoc* targets from *some* → *ad hoc* trials and *some* targets from *ad hoc* → *some* trials, and the number4 ↔ *ad hoc* panel consists of number4 and *ad hoc* targets from the number4 → *ad hoc* trials and the *ad hoc* → number4 trials respectively. The between-category panels show a difference between strong and weak trials, that is, between-category priming, although the effect is much smaller than that for the within-category primes.

We report three analyses. The first assessed whether EVAs can be primed at all, and if so, whether this effect occurs at the within and between-category levels. The model included a *within/between* factor that distinguished within-category trials from between-category trials (2 levels: pooled responses from *some* → *some*, number4 → number4, and *ad hoc* → *ad hoc*; and pooled responses from *some* → number4, number4 → *some*, *some* → *ad hoc*, *ad hoc* → *some*, number4 → *ad hoc*, *ad hoc* → number4), and a *prime* factor that distinguished strong primes from weak primes (2 levels: strong, weak). Table 1 reports statistical details of the analysis. A model using sum contrasts for both factors showed a significant effect of prime, $\beta = 0.56$, $p < .001$, such that a strong prime increased the rate of strong responding overall, a significant effect of within/between, $\beta = 0.126$, $p < .001$, such that the rate of strong responses was higher in between-category trials than within-category trials, and an interaction between the two, $\beta = -0.43$, $p < .001$, such that the effect of the prime was greater in the within-category trials. A model with the same structure but using treatment contrasts for the within/between factor and sum contrasts for the prime factor was used to investigate simple effects. This showed that significant priming occurred at the within-category level, $\beta = 0.99$, $p < .001$, and, independently, at the between category level, $\beta = 0.13$, $p < .001$. In short, we observed priming of EVAs at the within-category level and the between category level.

To assess these effects in more detail we broke down the data into within-category trials (Panels 1–3 of Fig. 2) and between-category trials (Panels 4–6 of Fig. 2) and conducted separate analyses on each. For the within-category analysis we tested a model with within-category group (3 levels: *some* → *some*, number4 → number4, *ad hoc* → *ad hoc*), and prime (2 levels: strong, weak) as factors, in which prime was coded with sum contrasts and target category with treatment contrasts (with *ad hoc* as reference). There was a significant effect of prime, $\beta = 1.24$, $p < .001$,

Table 1
Experiment 1 results.

		β	S.E.	Z	*p*-value
Overview	Prime * WithBet + (1 + Prime * WithBet |subject)				
	(Intercept)	−0.594	0.198	−2.991	.003
	Prime	0.563	0.034	16.342	<.001
	WithBet	0.126	0.029	4.284	<.001
	Prime: WithBet	−0.430	0.033	−13.177	<.001
Within simple	Prime	0.993	0.059	16.950	<.001
Between simple	Prime	0.133	0.033	4.082	<.001
Within detail	Prime * WithCat + (1 + Prime * WithCat |subject)				
	(Intercept)	−2.088	0.255	−8.185	<.001
	Prime	1.239	0.109	11.374	<.001
	WithCatNUM4	2.068	0.195	10.588	<.001
	WithCatSOME	1.823	0.157	11.598	<.001
	Prime: WithCatNUM4	0.174	0.166	1.046	.296
	Prime: WithCatSOME	−0.138	0.137	−1.007	.314
Between detail	Prime * BetCat + (1 + Prime * BetCat|subject)				
	(Intercept)	−0.691	0.204	−3.384	<.001
	Prime	0.145	0.058	2.509	.012
	BetCatSOMEADH	−0.054	0.089	−0.611	.540
	BetCatSOMENUM4	0.889	0.112	7.915	<.001
	Prime:BetCatSOMEADH	−0.069	0.079	−0.873	.383
	Prime:BetCat SOMENUM4	0.078	0.088	0.888	.374

Note. R-pseudo code shown in the first line of every section. *Prime* = priming factor (2 levels: strong, weak). *WithBet* = within/between factor (2 levels: within, between). *WithCat* = within expression category factor (3 levels: *some*, number4, *ad hoc*). *Betcat* = between expression category factor (3 levels: *some* ↔ number4, *some* ↔ *ad hoc*, number4 ↔ *ad hoc*).

consistent with the analysis above. There were also differences across categories in overall rates of strong responses. There were significantly more strong responses in the number4 category than the *ad hoc* category, $\beta = 2.07$, $p < .001$ and more in the *some* category than in *ad hoc* category, $\beta = 1.82$, $p < .001$, but there were no differences between number4 and *some* categories however, 0.25, $p = .14$. The interaction between prime and category was significant for number4 vs *some*, $\beta = 0.31$, $p = .03$ but not for number4 vs *ad hoc* nor for *some* vs *ad hoc*, $|b|$'s < 0.17, p's > .31.

To analyze the between-category data we formed three groups corresponding to the three possible between-category prime-target groups: *some* ↔ number4, *some* ↔ *ad hoc* and number4 ↔ *ad hoc*, corresponding to Panels 4–6 respectively in Fig. 2. Each group was the pooled responses from the two relevant target trials (i.e., groups were independent of direction). For example, the *some* ↔ number4 group consisted of *some* (prime) → number4 (target) trials and number4(prime) → *some* (target) trials. These groups formed three levels of one factor in the model, between category, and prime (2 levels: strong, weak) was another. Between category was coded with treatment contrast (number4 ↔ *ad hoc* as reference) and prime as sum contrasts. Replicating the results from the overview model, we found a significant effect of prime type, $\beta = 0.15$, $p < .001$, such that strong primes led to a greater rate of enrichment. There were no significant interactions of prime with between-category group, $|b|$'s < 0.15, p's > .082.

Discussion

Our findings reveal that enrichment can be primed: The decision about whether to enrich an expression was influenced by whether the expression was enriched on the preceding trial. Clearly, the mechanisms involved in computing enrichments are sensitive to recent activity. We were also able to identify different sorts of priming. In particular, we found within-category priming, between category priming, and greater within-category priming than between-category priming. The between-category priming effect illustrates that there are shared mechanisms across the EVA categories, yet the greater within-category priming result demonstrates that there is also some additional effect of EVA specific mechanisms.

The priming effect between *ad hoc* expressions and *some*/number4 helps eliminate an explanation for the *some* ↔ number4 priming effect. According to some authors (e.g., Horn, 1972), the alternatives for quantifier and number expressions are lexically defined (the alternatives are the stronger elements on the same semantic scale). A reasonable explanation for the *some* ↔ number4 priming effect was therefore that there was a special mechanism that retrieved the alternatives from memory, and that this mechanism was primed. For example, consider a strong *some* prime trial followed by a number4 target trial. The strong *some* interpretation would have meant activating the retrieval mechanism to obtain the alternative, *all*. If this mechanism had remained active into the target trial, it would have been more likely to retrieve the number alternative, thereby elevating the rate of enrichment. The *ad hoc* EVAs provide a test of this account. Since they do not have lexically defined alternatives (the alternative is defined entirely by the context), they could not share a lexical retrieval mechanism with *some*/numbers. Consequently, if a lexical retrieval mechanism were being primed, we would have expected reduced or non-existent priming between *some*/number 4 and the *ad hoc* EVAs (i.e., a between-category by prime interaction). Since we did not,[3] it is more parsimonious to assume that the same (non-lexical) mechanism is the source of the between-category priming effect.

Interestingly, while the size of the priming effects was similar across EVAs, the overall levels of enrichment differed. In particular, there was a much lower rate of enrichment for *ad hoc* EVAs than quantifiers or numbers. Thus, while the source of the between priming effect in Experiment 1 was a mechanism shared across *ad hoc*, numbers and quantifiers, there may nonetheless be differences in how the *ad hoc* and the other EVAs are computed, such as the computation of their alternatives (e.g., Katzir, 2007). We discuss this further in the General discussion.

Experiment 2

Experiment 2 was similar to Experiment 1 except that we used different EVA categories. The number sentences were *number4* sentences, just as before, and *number6* sentences, which involved *six*, such as, "There are six diamonds" (see Fig. 3). We also included *some* sentences. There were within-category trials that involved a prime and target from the same category, such as *some* → *some*, and between-category trials that involved a prime and target from different categories, such as *some* ↔ number4, just as there were in Experiment 1.

We had two aims. The first was to test an explanation for the within-category findings we observed in Experiment 1. Since the image for the weak prime picture had the same form as the image for the target picture, the priming effect could have been a consequence of participants being biased towards selecting a picture that was visually most similar to their previous selection. In a weak target trial, participants could select a picture that had an identical form to the picture that they selected in the weak prime trial, whereas in a strong trial, they could not (and so were obliged to choose the "Better Picture?" option). For example, consider the number4 trials (see Fig. 1). The weak prime consisted of a picture involving six symbols, together with the false picture, and the target consisted of another picture that contained six symbols, together with the "Better Picture?" option. Now, if participants adopted a strategy of choosing the picture most similar

[3] To test whether the failure to observe a significant interaction was due to experimental insensitivity, we computed the Bayes Factor for the interaction term. We conducted a Baysian repeated measures ANOVA on the between category data, with prime (strong *vs* weak), and between category group (*some* ↔ number4, *some* ↔ *ad hoc*, or *number4* ↔ *ad hoc*) as factors, using JASP (Love et al., 2015). We used the JZS prior (Rouder, Speckman, Sun, Morey, & Iverson, 2009), with r set *a priori* to the default value, $r = .707$. The Bayes Factor for the interaction term was 0.053, which is "substantial" evidence in favor of the null hypothesis (Dienes, 2011, 2014).

to their selection on the previous trial, they would select the six symbol picture following the weak prime (both had six symbols), and the "Better Picture?" option following the strong prime. This explanation could explain the within-category priming effect (but not the between category effect).

To test this hypothesis, we constructed our materials so that the weak number4 picture (six symbols) was the same as the strong number6 picture (six symbols). An image similarity account makes two predictions about responses to the number4 target trials: (1) participants should chose the weak picture (six symbols) more often after the number6 strong prime trial (six symbols), than after the number4 strong trial (four symbols) and (2) participants should chose the weak picture (six symbols) equally often after the weak number4 prime (six symbols) as after the strong number6 prime (six symbols).

Our second aim was to test a potential "lexical boost" to the priming effect. Research into syntactic priming has found that greater priming occurs when there is high lexical overlap between the prime and target sentences (e.g., Pickering & Branigan, 1998; Branigan et al., 2000; Cleland & Pickering, 2003). For example, Pickering and Branigan found a greater priming effect when prime and target used the same verb than when it did not. A similar effect in our study could clarify the mechanisms responsible for EVA priming effects, just as it has in the syntactic priming literature. We therefore compared trials in which prime and target had the same number ("four" → "four" and "six" → "six") with those that had different numbers ("four" → "six" and "six" → "four"). Greater priming in the same number trials compared to the different number trials would indicate a lexical boost.

Method

Participants

Ninety-six participants were recruited using Amazon Turk, all of whom declared English as their native language.

Materials

The number4 and *some* items were the same as those used in Experiment 1. Number6 items were constructed in similar way to number4 items except that six, nine, and four symbols were used for the strong, weak and false picture respectively. The filler sentences for number6 were *nine* sentences e.g., "There are nine diamonds." All other aspects of the design were the same as Experiment 1.

Results

Data treatment

We removed responses to incorrect primes, just as in the previous experiment. 487 target responses were removed out of a total of 6403. Of the 487, 142 were number4 targets, 176 were number6 targets and 173 *some* targets.

Analysis

Fig. 4 shows the proportion of strong responses to target trials in Experiment 2. Priming is shown by the

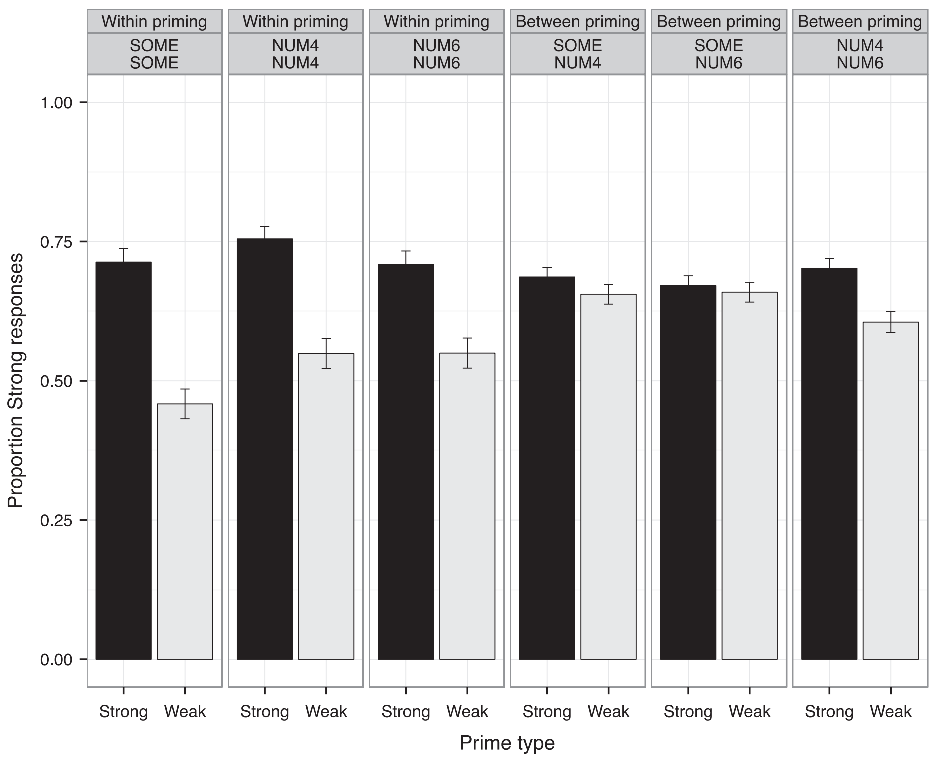

Fig. 4. *Experiment 2 results.* See Fig. 2 for explanation.

difference between the strong and weak primes within each panel. Just as in Experiment 1, there is a large within-category effect (Panels 1–3) and a smaller between-category priming effect (Panels 4–6).

We first report an overview analysis, just as we did in Experiment 1. The model included a *within/between* factor (2 levels: pooled responses from *some* → *some*, number4 → number4, and number6 → number6 *vs* pooled responses from *some* ↔ number4, *some* ↔ number6, number4 ↔ number6), and a *prime* factor (2 levels: strong, weak). Table 2 reports statistical details. There was a significant effect of prime, $\beta = 0.66$, $p < .001$, such that a strong prime increased the rate of strong responding, a significant effect of within/between, $\beta = 0.25$, $p < .001$, such rate of strong responding was greater in between-category groups than within category groups, and an interaction between the two, $\beta = -0.42$, $p < .001$, such that the effect of the prime was greater in the within-category trials. Simple effects analysis showed that significant priming occurred at the within-category level, $\beta = 1.08$, $p < .001$, and at the between category level, $\beta = 0.23$, $p < .001$. Our results therefore confirm the between and within priming observed in Experiment 1.

Next we consider the within-category effects in more detail. Here, we used only the within-category data (Panels 1–3 of Fig. 2). The model included within-category group (3 levels: number4 → number4, number6 → number6, and *some* → *some*), and prime (2 levels: strong, weak) as factors, with treatment contrasts for category group (number4 as reference) and sum contrasts for prime. There was a significant effect of prime, $\beta = 1.48$, $p < .001$, consistent with the overview analysis. There were also differences between the within categories in overall rates of strong responses. There were more strong responses in number4 than *some*, $\beta = 1.26$, $p < .001$, more strong responses in number6 than *some*, $\beta = 0.86$, $p = .0030$, and marginally more strong responses for number4 than number6, $\beta = -0.40$, $p = .058$. There were no significant interactions between prime and category, β's < 0.36, p's > .080.

The third analysis tested between-category priming. As in Experiment 1, the model included prime (2 levels: strong, weak) and between category group (3 levels: some ↔ number4, some ↔ number6, number4 ↔ number6) as factors and used only between-category data. Prime was coded with sum contrasts and between category group with treatment contrasts

Table 2
Experiment 2 results.

		β	S.E.	Z	p-value
Overview	Prime ∗ WithBet + (1 + Prime ∗ WithBet|subject)				
	(Intercept)	1.339	0.321	4.165	<.001
	Prime	0.657	0.053	12.457	<.001
	WithBet	0.246	0.049	5.056	<.001
	Prime: WithBet	−0.424	0.048	−8.739	<.001
Within simple	Prime	1.081	0.084	12.934	<.001
Between simple	Prime	0.231	0.057	4.066	<.001
Within detail	Prime ∗ WithCat + (1 + Prime ∗ WithCat|subject)				
	(Intercept)	2.021	0.439	4.605	<.001
	Prime	1.479	0.173	8.533	<.001
	WithCatNUM6	−0.402	0.212	−1.893	.058
	WithCatSOME	−1.260	0.262	−4.809	<.001
	Prime: WithCatNUM6	−0.357	0.204	−1.749	.080
	Prime: WithCatSOME	−0.327	0.210	−1.557	.120
Between detail	Prime ∗ BetCat + (1 + Prime ∗ BetCat|subject)				
	(Intercept)	1.894	0.447	4.240	<.001
	Prime	0.518	0.113	4.583	<.001
	BetCatSOMENUM4	−0.125	0.196	−0.637	.524
	BetCatSOMENUM6	−0.395	0.204	−1.930	.054
	Prime:BetCat SOMENUM4	−0.266	0.168	−1.583	.113
	Prime:BetCat SOMENUM6	−0.521	0.151	−3.451	<.001
Lexical boost	Prime ∗ Within + (1 + Prime ∗ Within|subject)				
	(Intercept)	1.996	0.436	4.573	<.001
	Prime	0.855	0.088	9.740	<.001
	WithBet	0.128	0.079	1.612	.107
	Prime: WithBet	−0.369	0.081	−4.546	<.001
Image similarity 1	PrimeCat + (1 + PrimeCat|subject)				
	(Intercept)	3.824	0.626	6.105	<.001
	PrimeCat	0.223	0.160	1.389	.165
Image similarity 2	PrimeCat + (1 + PrimeCat|subject)				
	(Intercept)	1.648	0.453	3.638	<.001
	PrimeCat	−0.895	0.157	−5.695	<.001

Note. R-pseudo code shown in the first line of every section. *Prime* = priming factor (2 levels: strong, weak). *WithBet* = within/between factor (2 levels: within, between). *WithCat* = within expression category factor (3 levels: *some*, number4, number6, but restricted to number4, number6 for lexical boost analysis). *Betcat* = between expression category factor (3 levels: *some* ↔ number4, *some* ↔ number6, number4 ↔ number6). *PrimeCat* = priming category factor (2 levels: number6, number4).

(number4 ↔ number6 as reference). We observed a significant between-category priming effect, $\beta = 0.52$, $p < .001$. However, we also observed significantly greater priming within number4 ↔ number6 than within some ↔ number6, $\beta = -0.52$, $p < .001$, and numerically greater priming within number4 ↔ number6 than some ↔ number4, $\beta = -0.27$, $p = .11$.[4] Furthermore, priming within number4 ↔ number6 was greater than priming within a combined quantifier-number group (some ↔ number4 trials and some ↔ number6 trials), as shown by the interaction between prime and quantifier-number, $\beta = -0.22$, $p < .0014$. Thus overall, priming within the number scale was greater than priming across the quantifier number scales, even when the lexical expressions were different in both groups. We discuss the significance of this in the discussion.

We also wanted to establish whether there was a lexical boost to the priming effect. We tested this by comparing the priming effect of the within-category number trials (number4 ↔ number4 combined with number6 ↔ number6) against the between-category number trials, (number4 ↔ number6) with within/between (2 levels: within, between) and prime (2 levels: strong, weak) as factors with sum contrasts. This analysis revealed a greater priming effect for within-category number trials than between category number trials, as shown by the significant interaction, $\beta = -0.37$, $p < .001$.

Finally, we assessed the image-similarity account of the within-category priming effect. There were two predictions, both relating to the number4 target. The first was that participants should have chosen the weak picture more often after the strong number6 prime trial than after the strong number4 trial. We therefore compared number4 target responses after a strong number6 compared to a strong number4 trial, with prime category (number6, number4) as a factor. In contrast to the image similarity hypothesis we failed to find to such an effect, $\beta = 0.22$, $p = 0.17$. The second prediction was that participants should have chosen the weak picture equally often after the weak number4 prime as after the strong number6 prime. Here again, we found evidence against the image-similarity account: for number4 target responses, the strong number6 prime led to significantly greater rates of enrichment than the weak number4 prime, $M = 0.70$ ($SD = 0.46$) *vs* $M = 0.55$ ($SD = 0.50$), $\beta = -0.90$, $p < .001$.

Discussion

Just as in Experiment 1, our findings show that enrichment can be primed. We found within-category priming, between-category priming, and greater within-category priming than between-category priming. The results of Experiment 2 also categorically rule out the image similarity explanation for the large within-category priming effect: we observed no difference between conditions where the image similarity account predicted there would be; and significant differences where it predicted there would not be.

We also found a lexical boost to the priming effect, consistent with the syntactic priming literature (e.g., Pickering & Branigan, 1998). There was a larger priming effect where numbers were the same (number4 ↔ number4, number6 ↔ number6) than when they were different (number4 ↔ number6). However, we also found that the between number effect (number4 ↔ number6) was greater than quantifier ↔ number effect (*some* ↔ number4, *some* ↔ number6). The latter result demonstrates the distinction between within-category priming that shares the same key lexical expression (e.g., *some* → *some*, number4 → number4 as in Experiment 1) and within-category priming that does not (e.g., number4 ↔ number6). There is a thus a lexical boost, and a distinct EVA boost.

Experiment 3

Sentences with plural nouns intuitively make reference to more than one object. To see this, consider the sentence (10a), which feels contradictory:

(10) John has chairs in his room; in fact he has exactly one.

The contradiction arises because "chairs" could be paraphrased as *more than one chair*, yet in the second clause refers to exactly one. However, consider (10b) or (10c), in which "chairs" is in a downward entailing context:

(10b) John doesn't have chairs in his room.
(10c) I don't think John has chairs in his room.

In these cases, "chairs" could not be paraphrased as *more than one chair*, since this would incorrectly allow the possibility that John had one chair in his room. Instead, the plural is best paraphrased as *at least one*, which excludes that John has exactly one chair in his room (or any other number). Hence in contexts like (10a), plurals seems to have a strong meaning, *more than one*, whereas in downward entailing contexts, like (10b) and (10c) they seem to have a weaker meaning, *at least one*. There is also psycholinguistic evidence that even in the absence of downward entailing environments the plural is underspecified for number (Patson & Ferreira, 2009; Patson, George, & Warren, 2014; Patson & Warren, 2011). For example, Patson et al. used a picture-matching paradigm to show that participants were just as fast to match a plural noun (e.g., *apples*) with a picture of a single object (an apple), as they were to match a plural noun with a picture of multiple

[4] Testing the effect of prime on each between-category group separately showed that there was a significant effect of prime on number4 ↔ number6, $\beta = 0.45$, $p < .001$, some ↔ number4, $\beta = .50$, $p < .001$, but not some ↔ number6, $\beta = 0.12$, $p = .35$. We cannot say for sure why some ↔ number6 did not demonstrate a significant priming effect but the reason might have been that the weak picture for number6 contained nine elements, which is the same as the number of elements in the *some* pictures. This could have caused interference that wasn't present for the some ↔ number4 group. Alternatively, the some ↔ number6 result might be a Type II error, since the interaction between prime and some ↔ number4, some ↔ number6 is not significant, $\beta = 0.25$, $p = .09$.

objects (multiple apples). Patson et al. concluded that the plural noun can generate a sentence representation that is unspecified for number. While there is no generally accepted explanation for the apparent paradox in plural meanings, Spector (2007) and others make an interesting argument that plurals have a basic meaning corresponding to *at least one*, and that the stronger, *more than one* interpretation is an enrichment derived using alternatives, that is, an EVA.[5]

In Experiment 3 we test whether numbers, quantifiers and plural enrichments prime each other. If they do, this would suggest that the *more than one* interpretation of plural morphology (intuitively the putative meaning) is derived using some of the same mechanisms as classical EVAs, consistent with the arguments of Spector (2007) and others. If they do not, there must be at least some differences between the two phenomena (since we have already shown that overlapping EVA mechanisms can be primed in general).

The general design was the same as that of Experiment 1 except that we used plural morphology items instead of *ad hoc* items. The plural items were of the form, "There are [symbol]s". The strong picture was three symbols consistent with the predicate, the weak picture one symbol consistent with the predicate, and the false picture one symbol, inconsistent with the predicate. Fig. 3 shows examples.

Method

Participants

One hundred participants were recruited using Amazon Turk. Of these, six were removed because they did not declare English as their native language.

Materials

The number4 and *some* items were the same as those used in Experiment 1. Instead of number6 items we included plural items.

The experimental plural sentences were "There are [symbol]s." Weak plural pictures contained a single symbol and strong plural pictures contained three symbols. False plural pictures contained a single symbol that did not match the predicate. Strong and weak prime trials were constructed in the same way as strong and weak *some* and number4 trials. Fig. 3 shows examples.

Filler plural trials involved the alternative, just as with the *some* and number4 trials. This was implemented as, "There is a single [symbol]." The construction of these items followed the same three formats as for the *some* and number4 trials described in Experiment 1. The remainder of the design and procedure was identical to Experiment 1.

Results

Data treatment

278 trials out of 6825 responses were removed. Of these, 107 were number4 trials, 38 were plurals, and 133 were *some* trials.

Analysis

Fig. 5 shows the data from Experiment 3. As with the previous experiments, there was a large difference between strong and weak primes for within-category trials. There is also a large between-category priming effect between *some* and number4 but a much smaller between-category effect involving the plurals.

The overview analysis revealed a significant effect of prime, $\beta = 0.65$, $p < .001$, of within/between, $\beta = 0.30$, $p < .001$, and the interaction between them, $\beta = -0.62$, $p < .001$ (see Table 3 for the statistical details). Similarly, the simple effects analysis showed a significant within-category priming effect, $\beta = 1.27$, $p < .001$. However, there was no significant between-category priming effect, $\beta = 0.033$, $p = 0.50$. This would be expected if there were no priming between plural items and quantifier/numbers, as we probe in more detail below.

The within-category priming results were similar to other experiments. The effect of prime was significant overall, $\beta = 1.27$, $p < .001$, and the rate of enrichment differed across categories such that the plurals were enriched more often than number4, $\beta = 0.68$, $p < .001$, and marginally more than *some*, $\beta = 0.51$, $p = .012$, but *some* and number4 did not differ, $\beta = 0.17$, $p = .44$. There was also an interaction such that the effect of the prime was larger on the plural items than the number4 items, $\beta = 0.51$, $p < .001$, but not larger on the plural than the *some* items, $\beta = 0.29$, $p = .12$, nor was there a difference between *some* and number4 categories, $\beta = 0.22$, $p = .25$.

The between category analysis used between category group (*some* ↔ number4, plural ↔ number4 and plural ↔ *some*) and prime (strong, weak) as factors in the model. This analysis revealed no significant effect of prime, $\beta = -0.10$, $p = .23$, consistent with the overview model, but there were significant interactions between prime and between-category groups. Treatment coding of between category group revealed that the effect of prime was greater on *some* ↔ number4 than on plural ↔ number4, $\beta = 0.30$, $p = .0080$, and greater on some ↔ number4 than plural ↔ *some*, $\beta = 0.28$, $p = .017$, while there was no difference between the prime on *some* ↔ plural and on plural ↔ number4, $\beta = 0.028$, $p = .81$. In short, the effects of the prime were larger when priming did not involve the plural trials than when it did. As a further test of this we combined the plural trials (plural ↔ *some* and plural ↔ number) and compared them with non-plural trials (*some* ↔ number4), in a model with plural and prime as factors (summed contrasts). This showed the expected interaction of prime by plural, $\beta = 0.16$, $p = .0020$. Simple effects analysis showed a significant effect of prime on non-plurals (*some* ↔ number4), $\beta = 1.09$, $p < .001$, but no

[5] In this particular case, one needs to assume that a plural such as "chairs" has as an alternative meaning something like "a unique chair", so that its negation (*not a unique chair*) yields the plural meaning. One challenge is to explain where this complex alternative comes from. We need not dive into this question here, but note that current accounts typically assume that this alternative is itself obtained as the enriched meaning of another, simpler expression (a singular DP). The relevance of this is that it makes EVAs associated with plurals special in at least one respect: they show up as a kind of second order EVA (just like free choice, see e.g., Fox, 2006).

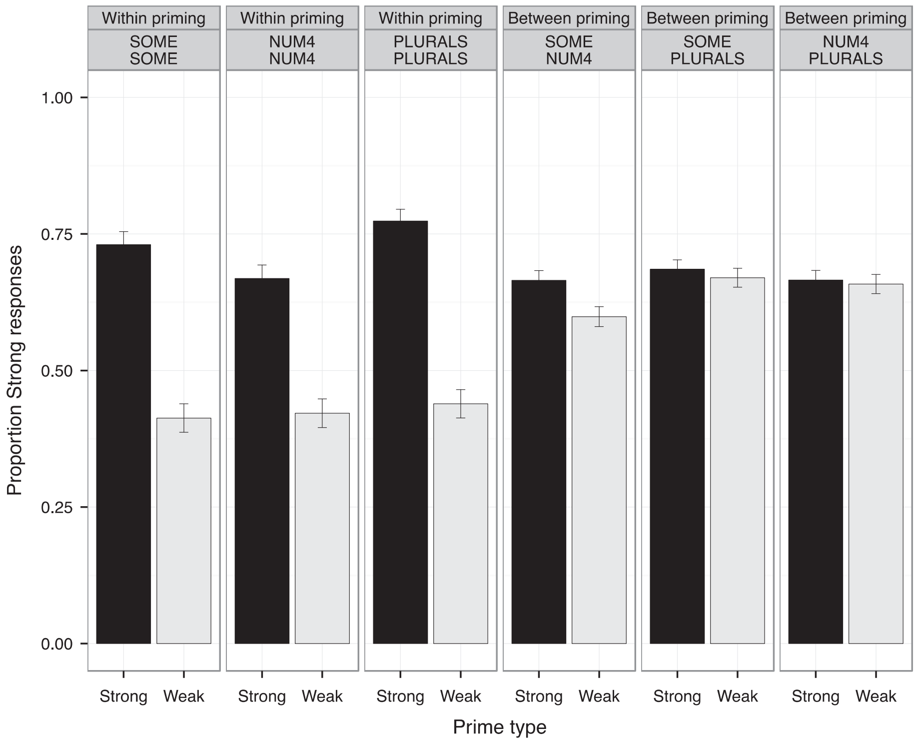

Fig. 5. *Experiment 3 results.* See Fig. 2 for explanation.

Table 3
Experiment 3 results.

		β	S.E.	Z	p-value
Overview	Prime * WithBet + (1 + Prime * WithBet|subject)				
	(Intercept)	0.917	0.267	3.433	<.001
	Prime	0.650	0.048	13.523	<.001
	WithBet	0.295	0.041	7.176	<.001
	Prime: WithBet	−0.618	0.054	−11.484	<.001
Within simple	Prime	1.267	0.089	14.245	<.001
Between simple	Prime	0.033	0.049	0.668	.504
Within detail	Prime * WithCat + (1 + Prime * WithCat|subject)				
	(Intercept)	0.445	0.324	1.375	.169
	Prime	1.174	0.164	7.141	<.001
	WithCatPLURALS	0.679	0.200	3.392	<.001
	WithCatSOME	0.171	0.220	0.777	.437
	Prime: WithCatPLURALS	0.506	0.177	2.863	.004
	Prime: WithCatSOME	0.223	0.192	1.160	.246
Between detail	Prime * BetCat + (1 + Prime * BetCat|subject)				
	(Intercept)	1.408	0.311	4.534	<.001
	Prime	−0.104	0.087	−1.196	.232
	BetCatSOMENUM4	−0.325	0.145	−2.237	.025
	BetCatSOMEPURALS	0.260	0.174	1.490	.136
	Prime:BetCat SOMENUM4	0.303	0.114	2.656	.008
	Prime:BetCatSOMEPURALS	0.028	0.117	0.243	.808
Plurals vs non-plurals	Prime * BetPls + (1 + Prime * BetPls|subject)				
	(Intercept)	1.278	0.298	4.284	<.001
	Prime	0.047	0.054	0.868	.385
	BetPls	−0.192	0.063	−3.047	.002
	Prime:BetPls	0.157	0.051	3.085	.002
Plural simple	Prime	−0.110	0.069	−1.587	.113
Non-plural simple	Prime	1.086	0.301	3.610	<.001

Note. R-pseudo code shown in the first line of every section. *Prime* = priming factor (2 levels: strong, weak). *WithBet* = within/between factor (2 levels: within, between). *WithCat* = within expression category factor (3 levels: *some*, number4, plurals). *Betcat* = between expression category factor (3 levels: *some* ↔ number4, *some* ↔ plurals, number4 ↔ plurals). BetPls = plurals factor (2 levels: plural combinations, non-plural combinations).

effect of prime on plurals (plural ↔ *some* and plural ↔ number4), $\beta = -0.11$, $p = .11$. Overall we find no evidence that plurals and *some*/number prime each other.

Discussion

The primary goal of this experiment was to test whether the numbers and the quantifiers share enrichment mechanisms with plural morphology. Our evidence suggests that there are at least some mechanisms used by *some* and the numbers that are not shared with plurals. While we found as much within-category priming for the plural items as *some* and the numbers (indeed more than for the numbers), we found significantly less between-category priming between plurals and numbers/quantifiers than between quantifiers and numbers. We also observed robust between category priming between quantifiers and numbers, but none between quantifiers/numbers and plurals. These findings argue against equating the derivation of plural morphology interpretation with the derivation of classical EVAs (we consider this conclusion in more detail in the General discussion).

The results of Experiment 3 nonetheless provide information about what sort of mechanism is being primed between the numbers and quantifiers (and indirectly the *ad hocs*). In particular, it eliminates the possibility that the only source of between-category priming was that participants were being primed to derive the most informative interpretation of the sentence (or, equivalently, the most exact, or precise, interpretation). Perhaps people have a general bias towards weak interpretations (which are generally more likely to be true) and the prime gave them sufficient confidence in the speaker's knowledge to select the strong interpretation. This is plausible because there is independent evidence that children, in particular, favor interpretations that make a statement true in a context (Crain & Thornton, 1998) and that people are generally sensitive to the informativeness of sentences independent of whether they derive an enrichment (Katsos & Bishop, 2011). However, in our experiment, the strong and weak interpretations of the plural items also formed an informativeness scale. If the between-category priming results between the numbers and quantifiers were due to priming of informativeness, we should have observed the same level of priming between numbers and quantifiers as between numbers/quantifiers and the plurals, which we did not. Instead, we must have been priming something specific to EVAs, rather than to informativeness in general.

Combined analysis of Experiments 1–3

We tested priming between *some* and *four* in all three experiments. Since sampling issues are likely to be minimal here (we used repeated measures manipulations throughout) we combined the experiments to create a larger (and more powerful) data set. We then used this data to investigate two further questions about how participants were completing our task.

The first was whether the direction of priming was important, that is, whether *some* → number4 results in different priming effects than number4 → *some*. Despite similarities, some differences remain between these two types of EVA: the EVA with numbers is harder to cancel and occurs more easily in a wider variety of linguistic environments than the EVA with *some* does. These differences may be tied to the way the alternatives are retrieved, such that retrieving the alternative of some would be a strictly more complex process than retrieving those for *four*, for instance. If this were the case, and if the between priming effect occurs at the level of this alternative retrieval process, then we could expect the priming to be asymmetrical.

The second question we asked was whether the priming effect differed between the first and the second half of our experiments. This could tell us how task dependent our effects were. For example, if the effects existed only in the second half, one might argue that participants needed to be "taught" (through repeated exposure) to understand one or other of the sentence meanings.

Analysis and discussion

Combining all three experiments resulted in a data set with $N = 377$ participants. The results are shown in Fig. 6. We first conducted an overview analysis with prime (weak, strong) and within/between as factors, as in previous analyses. This revealed the expected significant effects of prime, $\beta = 0.60$, $p < .001$, within/between, $\beta = 0.20$, $p < .001$, and the interaction, $\beta = -0.48$, $p < .001$ (Table 4 shows the statistical details). Simple effects analysis showed a significant effect of prime at the within category level, $\beta = 1.08$, $p < .001$, and the between category level, $\beta = 0.12$, $p < .001$, replicating our previous findings.

Next we assessed the directional effects in the between category data. We constructed a model with prime (strong, weak) and direction (number4 → *some*, *some* → number4) as fixed effects, using sum coding, and applied the model to the between-category data. This showed a significant effect of prime, $\beta = 0.30$, $p < .001$, no effect of direction, $\beta = -.060$, $p = .43$, and no interaction, $\beta = .073$, $p = .13$. Furthermore, the simple effect of prime was significant for number4 → *some* trials, $\beta = 0.22$, $p < .001$, and for *some* → number4 trials, $\beta = 0.35$, $p < .001$. In short there was no evidence that *some* → number4 priming was larger than number4 → *some* priming, and good evidence that priming occurred both directions, contrary the alternatives retrieval explanation of the between-category priming result.

Finally, we assessed order effects by comparing the first half of the experiment with the second half. We used a model similar to the detailed within-category and between-category analyses shown in the individual experiment analyses. For within-category priming, we used prime (2 levels: strong, weak), within-category group (2 levels: *some*, number4) and half (2 levels: first half, second half) as sum contrasted factors. There was a significant effect of prime, $\beta = 1.43$, $p < .001$, but also an interaction between prime and half, $\beta = .21$, $p < .001$. However, analysis of prime using data restricted to the second half only showed a significant effect of prime, $\beta = 1.29$, $p < .001$, as it did for data restricted to the first half only, $\beta = 1.49$, $p < .001$. Thus while there was a slightly smaller effect of prime in the second half, effects were present in both. This

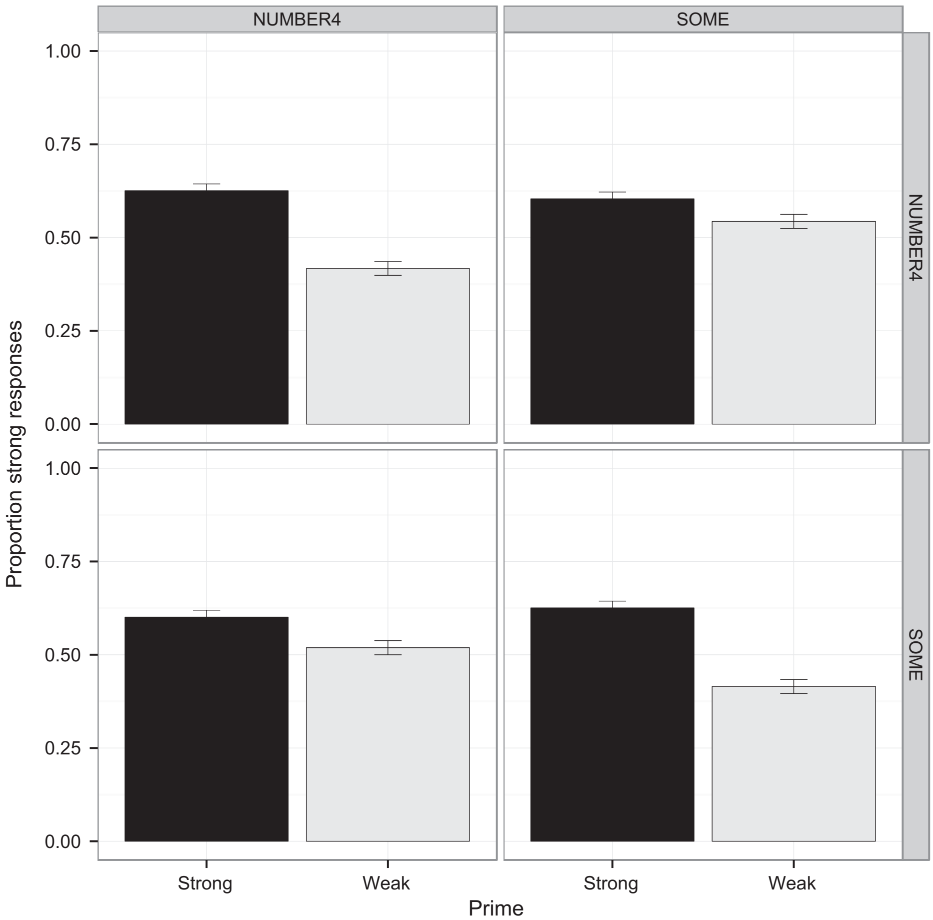

Fig. 6. *Combined analysis across Experiments 1, 2 and 3.* The strong and weak bars refer to the value of the prime, as in previous experiments. The panel rows of the figure refer to the prime category (*some* or number4), and the panel columns the target category (*some* or number4). Thus, the upper to lower diagonal shows within-expression trials and the lower to upper diagonal show between-expression trials.

suggests that the priming effect is strong *a priori* but reduced when we try to continuously alternate between opposite priming directions.

For between-category data, we used prime (2 levels: strong, weak), between-category group (2 levels: *some* ↔ number4, number4 ↔ *some*) and half (2 levels: first half, second half) as sum contrasted factors. The effect of prime was significant, $\beta = 0.34$, $p < .001$, and there was no interaction with half, $\beta = -.06$, $p = .26$. Thus there was no evidence that the effect differed by half.

General discussion

Our goal was to investigate the interaction between shared and distinct EVA mechanisms: are there core mechanisms shared by all EVAs, or does enrichment take place using distinct mechanisms for each linguistic category? Our results indicate that there are shared EVA mechanisms for quantifiers, numbers and *ad hoc* inferences and that the shared mechanisms are at least partially distinct from the mechanisms used in plural morphology.

Priming of EVAs

We have identified two sorts of EVA priming: within-category and between-category priming. Neither form has been reported in the literature previously. Here we discuss explanations for our effects from the perspective of the core account of enrichment (described in the Introduction).

Between-category EVA priming

The core account assumes that alternatives are constructed, or retrieved, and passed to the processor to be negated. If the speaker is sufficiently knowledgeable, the negated alternative is combined with the basic meaning to form the enriched meaning. Since these mechanisms are independent of specific EVAs, priming of each could, in theory, explain the between-category priming effect. Here we discuss these in more detail.

One possibility is that we were priming perceptions of how knowledgeable the speaker appears (or judgments about the mental state of the speaker). The idea would

Table 4
Combined analysis.

		β	S.E.	Z	p-value
Overview	Prime * WithBet + (1 + Prime * WithBet|subject)				
	(Intercept)	0.259	0.150	1.728	0.084
	Prime	0.601	0.025	24.382	<.001
	WithBet	0.196	0.022	9.078	<.001
	Prime:WithBet	−0.479	0.024	−20.024	<.001
Within simple	Prime	1.080	0.042	25.524	<.001
Between simple	Prime	0.120	0.024	5.064	<.001
Between details	Prime * Direction + (1 + Prime * Direction|subject)				
	(Intercept)	0.984	0.208	4.727	<.001
	Prime	0.294	0.048	6.099	<.001
	Direction	−0.060	0.075	−0.790	.430
	Prime:Direction	0.073	0.048	1.522	.128
Some → number4	Prime	0.354	0.069	5.104	<.001
Number4 → some	Prime	0.221	0.065	3.412	<.001
Within by half	Prime * WithCat * Half + (1 + Prime * WithCat * Half | subject)				
	(Intercept)	0.473	0.192	2.462	.014
	Prime	1.425	0.077	18.601	<.001
	WithCat	0.235	0.073	3.208	.001
	Half	0.145	0.062	2.355	.019
	Prime: WithCat	0.177	0.063	2.806	.005
	Prime:Half	0.212	0.056	3.787	<.001
	WithCat:Half	0.108	0.060	1.789	.074
	Prime:WithCat:Half	0.079	0.057	1.394	.163
Half 1 only	(Intercept)	0.531	0.188	2.829	.005
	Prime	1.488	0.086	17.222	<.001
	WithCat	0.183	0.091	2.012	.044
	Prime: WithCat	0.063	0.080	0.790	.430
Half 2 only	(Intercept)	0.423	0.211	2.001	.045
	Prime	1.295	0.091	14.260	<.001
	WithCat	0.153	0.082	1.879	.060
	Prime: WithCat	0.134	0.074	1.817	.069
Between by half	Prime * BetCat * Half + (1 + Prime * BetCat * Half |subject)				
	(Intercept)	1.096	0.240	4.557	<.001
	Prime	0.344	0.056	6.182	<.001
	BetCat	−0.027	0.086	−0.315	.753
	Half	−0.010	0.084	−0.118	.906
	Prime: BetCat	0.065	0.055	1.177	.239
	Prime:Half	−0.060	0.054	−1.122	.262
	BetCat:Half	−0.103	0.063	−1.628	.104

Note. R-pseudo code shown in the first line of every section. Factor expressions as in Table 1 except that levels are restricted to *some* and number4 and their combinations. *Direction* = prime-category → target factor (2 levels: *some* → number4, number4 → *some*). Half = experiment half factor (2 levels: first half, second half).

be that after a strong prime trial, the participant believes that the speaker is knowledgeable enough to have used the alternative and consequently derives the enrichment in the target trial, whereas after a weak prime trial, the participant is not confident about the speaker's knowledge, and so does not derive the enrichment in the target trial. However, there are several reasons why this seems implausible. The first is that our manipulation was a repeated measures design and there is little reason for participants to believe that the speaker on strong trials was different to the speaker on weak trials. We did not present the speaker differently across prime trials by, for example, creating one female speaker for strong trials and one male speaker for weak trials. The strong and weak primes were presented identically and in sequence, so there was no reason for the participant to distinguish between them. The second is that it would have been difficult for the participant to determine that the strong prime speaker was knowledgeable and that the weak speaker was not, even if they believed that there were multiple speakers. We did not use a cover story that manipulated the knowledge of the speaker (cf. Bergen & Grodner, 2012), nor did we vary speaker reliability (cf. Grodner & Sedivy, 2011). Instead, the participant would have had to engage in a form of backward Gricean reasoning, in which they reasoned that because the speaker was obliging them to derive an enriched interpretation in the strong prime trial, the speaker must be knowledgeable, whereas because the speaker did not oblige them to derive an enrichment for the weak trials, the speaker must not be knowledgeable.

A more plausible explanation is that we were priming the search for alternatives. On this account, there is a mechanism that can be primed to retrieve (or construct) relevant alternatives. Consider how this could explain our findings. On a strong prime trial the processor would be unable to provide a categorical response using a weak interpretation of the EVA expression (both responses options are consistent with the weak interpretation). This causes the participant to process the sentence more deeply, ultimately retrieving alternatives to what the speaker could have said and deriving the enrichment. On a weak prime trial, the weak interpretation provides a satisfactory response and there is no need to process the sentence more deeply. If the search mechanism has activation levels that retain their value across the inter-trial-interval, activation levels would be higher following a strong prime than a weak prime, and the enrichment would be more likely to arise. Note that this account does not assume that the alternatives *per se* are being primed (e.g., *all*), only that the *search* for alternatives is primed.

Finally, our effects could be explained by the mechanism that negates the alternative and combines it with the basic meaning (a usage mechanism). During a strong trial, in which enrichment occurs, the usage mechanism would be activated, whereas in a weak trial, where enrichment does not occur, the usage mechanism would not be activated. Consequently if the activation levels of the usage mechanism remain active across trials then the alternative would more likely be negated after a strong trial than after a weak trial. Under this account, the usage mechanism is a distinct mechanism that operates independently of the general salience of alternatives.

The status of the search and usage mechanisms has not been considered much in the literature (there has been no empirical need) but nor is either option inconsistent with prevailing accounts. In terms of general plausibility, continuously searching for alternatives is intuitively a costly activity, especially since those alternatives would not be used often, and having a mechanism that is triggered only on certain occasions would reduce the cost. Conversely, however, the search may not incur a cost if it is part of the general process of language comprehension. For instance, the search may be part of incremental prediction (e.g., Altmann & Kamide, 1999; DeLong, Urbach, & Kutas, 2005; Kutas & Hillyard, 1984), which presumably occurs all the time, and which would obviate the need for a trigger mechanism. Furthermore, the lack of between-category priming with plural items does not distinguish between the two accounts. If a search mechanism were being primed, the lack of an effect with plurals could be explained by a specialized plural search mechanism, and if a usage mechanism were being primed, by a specialized plural usage mechanism (or that alternatives were not used to derive plurals at all). In short, there is little to choose between the priming of a usage mechanism and a search mechanism based on our data or from previous research.

Within-category priming

Priming of the search and usage mechanisms explain well the between-category priming effects but they offer no explanation for why we observed a greater within-category priming effect than a between-category priming effect. If all that was being primed was the use of alternatives, say, we should have observed the same magnitude of within-category priming effect as between-category priming effect, which we did not. Additional mechanisms are therefore required to explain the elevated within-category effect. These are discussed below.

It is possible that participants were primed to accept different degrees of informativity (or precision). As we suggested in Experiment 3, people might have a bias towards weak interpretations (which are generally more likely to be true) and the strong prime gave them sufficient confidence in the speaker's knowledge to select the strong interpretation; or perhaps they are generally accepting of imprecise interpretations (the weak target) unless they have a reason to think that the speaker is being particularly exact (in which case they reject the weak target in favor of the better picture option). While the results of Experiment 3 eliminate this explanation of between-category priming, they do not do so for within-category priming. However, many of the arguments against priming of speaker beliefs that we made in the Between-category priming section are equally applicable here. Priming people to accept different degrees of informativeness requires them to judge that the speaker has different communicative requirements across trials (within the same experiment), or that there are different speakers across trials. Neither possibility seems plausible given that strong and weak prime trials were identical in presentation. Furthermore, if we were priming acceptance of informativity across within-category trials, we should have observed an equally large effect in the between-category priming trials, since in both conditions the weak interpretation was less informative than the strong interpretation. That we observed a much larger within-category effect than between-category effect argues against this explanation.

An alternative is that within-category priming could be explained by links between the trigger expressions and the derivation mechanisms. In the case of *some*, for example, there might be a link between *some* and the usage mechanism, so that repeated application of the usage mechanism while *some* is activated leads to a strengthening of the link between them. In other words within-category priming is lexical in nature. While there might be a lexical contribution to the within-category priming effect we doubt that it entirely explains the result, however. With *some*, the explanation seems plausible but with the other expressions we used, the numbers and the *ad hoc* EVAs, it is far less so. With numbers we observed a larger priming effect when the same number was used across prime and target (e.g., *number4* → *number4*) compared to when different numbers were used (*number4* ↔ *number6*). This is indeed evidence of a lexical priming effect. However, we also observed greater priming when different numbers were used (*number4* ↔ *number6*) than when a number and *some* was used (*some* ↔ number4, *some* ↔ number6). This suggests that in addition to a lexical effect, there is a within-EVA effect that is non-lexical in nature. Finally, while the *ad hoc* items involved the same words in each case, "There is a [symbol]", it seems unlikely that any one of them had pre-existing links to a usage mechanism since each of the words occur in many linguistic environments that do not generate enrichments (indeed, this is what makes them *ad hoc* EVAs). One would have to assume that "There is a" became lexicalized during the experiment and developed a link to the usage mechanism, which could then be primed.

A more likely explanation is that the primes altered the saliency of specific alternatives. A strong prime trial would force the participant to consider the alternatives whereas a weak prime trial would not. In the case of *some,* for example, *all* would be more salient after a strong prime than a weak prime. (For other categories, such as the *ad hoc* trials, the activated alternative would have to have a more abstract form). If the alternative remained highly active across trials, and if a salient alternative facilitated enrichment, more enrichment would be expected after strong primes than weak primes. The same effect would not be present in between-category priming trials because the alternatives would be different across trials.

Summary

Our results illustrate that enrichment depends on whether the immediate context includes prior enrichment. We have suggested different ways in which this can be explained within the framework of the core account, and expressed preferences as to the more plausible among them. In particular, we argue that (1) enrichment raises the saliency of alternatives, which leads to more enrichment and (2) enrichment primes either a search mechanism or a usage mechanism (or both), which also leads to enrichment.

Implications for theories of individual EVAs

Our experiments were not intended to address individual theories about enrichment but they nonetheless have implications for several debates in the literature. We discuss these below.

EVA vs exact accounts of numbers

Recall from the introduction that there are divergent theories of number representation. According to EVA accounts (Gazdar, 1979; Horn 1989; Levinson, 1983), numbers have a weak semantic representation (e.g., "four of the symbols are squares" means *at least four of the symbols are squares*) combined with an enrichment derivation of the strong interpretation, whereas exact accounts (e.g., Breheny, 2008; Geurts, 2006), favor a strong semantic meaning (e.g., "four of the symbols are squares" means *exactly four of the symbols are squares*) with a secondary weak meaning. Our data most straightforwardly provide evidence consistent with an EVA account of number expressions: numbers and quantifiers both have the weak meaning as basic, and both have the strong meaning derived using alternatives. Since they have the same sort of semantic representation and derivation mechanism, activation from one EVA category in a prime should lead to activation in the other EVA category in the target, exactly as we observed. Under the exact view, in contrast, numbers and quantifiers have a different sort of semantic representation and a different derivation. There is therefore no straightforward reason why activation of the mechanisms in one category would lead to activation of the mechanisms in the other category.

One way of reconciling our results with previous studies is to argue that observed differences between numbers and other types of EVAs (e.g., Guasti et al., 2005; Huang et al., 2013; Marty, Chemla, & Spector, 2013; Musolino, 2004; Papafragou & Musolino, 2003) arise for peripheral reasons. For instance, it could be that number enrichment is an easier task than *some* enrichment for independent reasons (just like apparently irrelevant factors can affect the difficulty of some reasoning tasks e.g., Johnson-Laird & Bara, 1984; Newstead, Pollard, & Allen, 1992). As a consequence, exact readings of numerals may be accessed more easily and more broadly, but yet through the same mechanisms as other EVAs. In the long run, one could hope to reconcile the processing differences and the priming similarities more ambitiously, possibly finding for instance that processing delays are caused by computation difficulties under certain circumstances (present for *some* but not for numbers), and priming effects are based on the activation of a common set of mechanisms, independently of their difficulties.

Plural morphology

We did not observe a priming effect between plurals and other EVAs. Given the consistency of between-priming effects throughout the other categories of EVAs, the resistance of plurals to enter into these priming effects challenges recent accounts that attempt to unify plural morphology with other triggers of EVAs (e.g., Spector, 2007). We see two ways to go about this new finding. First, one may abandon the EVA account of plurals. Second, less trivially, one may like to capitalize on some aspects of the EVA accounts of plurals that make them different from others and investigate whether these peculiarities play a key role in priming. There are several possibilities. The first is that plurals might use the same core mechanisms as *some*/numbers (derivation of the alternative and its negation) but to different degrees: Plurals may require the repeated application of the core mechanisms whereas *some*/numbers may require only a single application (e.g., in Spector (2007), plurals rely on *strengthened* alternatives, which themselves require deriving enriched meanings, whereas classical EVAs rely on simple alternatives, which do not). Plurals might therefore be more difficult to prime than classical EVAs but use the same processes. Another possibility is that the alternatives for plurals might be derived in a different way to the alternatives for *some*/numbers. For example, plural alternatives may involve deletion of the plural morpheme, rather than replacement of a lexical item, which happens to make a difference according to Katzir's (2007) view of alternatives. If the between-category priming effect in *some*/numbers arises because the mechanism that derives the alternatives is primed, no priming effect would be predicted if plurals use a different method of generating the alternative to *some*/numbers (even though plurals might share some of the remaining EVA mechanisms).

Overall then, our failure to find priming effects between plurals and the classical EVAs shows that the two phenomena do not share exactly the same set of mechanisms for their derivation, but they may nonetheless share some of the them, which are not the locus of priming effects. In the future, a better understanding of the locus of the priming effects could help locate the similarities and differences between plurals and other forms of EVAs. For instance, testing priming of plurals and *some*/number EVA with other appropriately chosen phenomena (e.g., free choice enrichments, which are also hypothesized to use strengthened alternatives) would clarify whether plurals use any of the same mechanisms as classical EVAs.

Ad hoc EVAs

Hirschberg (1991) and others have suggested that *ad hoc* enrichments arise by negating alternatives and combining them with the basic meaning, much like quantifiers and numbers. The between-category priming effects that we observed largely support this conclusion. Nonetheless, *ad hoc* EVAs were different in one respect: they exhibited substantially lower rates of overall enrichment than the other EVAs. Presumably the alternatives for the *ad hoc* EVAs were less available than those for the other EVAs.

We suggest that this was caused by the relative complexity of the *ad hoc* alternatives. While the alternatives for quantifiers and numbers were formed by substituting a single term for another (e.g., "all" for "some"), those for *ad hoc* EVAs required deriving extra material with meaningful content. For example, the alternative to "There is a diamond," was the conjunctive expression, "There is a diamond and a square." The extra material presumably requires more work from the processor relative to a simple substitution. Consequently the processor would fail a greater number of times in retrieving *ad hoc* alternatives compared to quantifier or number alternatives, and thus the enrichment would also fail on greater number of occasions. Such an account would be consistent with the views of theorists who advocate a role for the source of lexical material available for the replacement (e.g., again, Katzir, 2007) but the psychological claim – that the complexity of the alternative at least partially determines the rate of enrichment – remains to be tested.

Enrichment and structural priming

We have so far presented enrichment priming as being quite different to other sorts of structural priming reported in the literature. However, there are similarities and differences between the two that are profitable to consider.

Mechanisms vs representations

The structural priming literature typically refers to structures and representations, whereas we have described the standard account in terms of mechanisms for alternative retrieval and manipulation. While these two approaches appear to be quite different, EVAs could be seen as meaning-based representations (something like an EVA logical form) that can be primed (rather like Raffray and Pickering's (2010), account of scopal ambiguity priming). For example, the processor might represent two EVA structures, one for the strong interpretation and one for the weak interpretation:

(R1) X but not Alt[X]
(R2) X (and either Alt[X] or not Alt[X])

Where X corresponds to the EVA trigger and Alt[X] the alternatives to X. (In the recent grammatical version of the core account, as in Chierchia et al. (2012), the distinction may be more transparently linked to representational differences: R1 would correspond to a parse of the sentence containing the silent exhaustification operator O, and R2 would be a parse without this operator). R1 and R2 would be linked to EVA trigger expressions, such as *some* and the numbers, or particular structures, such as *There is a* [symbol] (the *ad hoc* structure), so that processing of these expressions would trigger the activation of both representations. The interaction between within and between-category priming would arise because during within-category trials, the appropriate enrichment representation retains elevated activation from the previous trial, and also receives a boost from the elevated activation on the trigger link, whereas during between-category trials, it is only the activation on the representation that contributes to enrichment (similar to Pickering and Branigan's (1998), explanation of lexical boost and syntactic priming).

Because priming effects in syntax have been linked to a representational view of syntax, one may thus investigate what would be the properties of a similar representational view of the priming effects we find for EVAs. By analogy with the syntactic priming literature, it may assume a level of representation intermediate between individual words and whole sentences at which we can recognize the phenomenon: in this view it is the activation of partial chunks, stored as such in the lexicon, which generate the priming effects. Here, for instance, these chunks could involve a combination of an exhaustivity operator (e.g., van Rooij and Schultz, 2004; Chierchia, 2004) and an enrichment trigger. Also, the representational view of priming typically assumes that there is no default as to which possible chunk is activated, in common with constraint-based models of language (Elman, Hare, & McRae, 2004; McRae, Spivey-Knowlton, & Tanenhaus, 1998). Whether these properties accurately reflect the behavior of EVAs remains to be seen (e.g., Tomlinson et al. (2013), present data against a straightforward constraint-based model of EVAs, although Degen and Tanenhaus (2015), show evidence in favor) but presenting EVAs as representations at least allows the similarities (and differences) between EVA priming and other forms of priming to be more apparent. In the much longer run, it could help nurture the debate about the semantic/pragmatic status of EVAs.

Inverse preference effect

In the syntactic priming literature there is evidence that the less preferred syntactic construction is a more effective prime that the more preferred structure (the *inverse preference effect*; Hartsuiker & Kolk, 1998; Hartsuiker, Kolk, & Huiskamp, 1990; Hartsuiker & Westenberg, 2000; Scheepers, 2003). In other words, structures that are more surprising are stronger primes than those that are less surprising. For example, English passives, which are less frequent than active structures, produce strong priming effects, but active passives do not (e.g., Bock, 1986). The inverse preference effect has been used to argue that priming is based on prediction error (Chang, Dell, & Bock, 2006; Fine & Jaeger, 2013; Jaeger & Snider, 2013). The basic idea is that people adjust their expectations about upcoming linguistic structure by minimizing the error between the predicted and observed linguistic structure. Since dispreferred structures result in larger prediction error, the prediction that the dispreferred structures will occur in subsequent trials will be adjusted more than the prediction associated with the preferred structures. Furthermore, because learning by minimization of prediction error is assumed to be implicit (rather than episodic), the inverse preference effect has also been used as evidence that priming is an implicit learning effect (Pickering & Ferreira, 2008). Do the priming effects we observe also exhibit the inverse preference effect?

Establishing this is somewhat complicated because we do not have *a priori* neutral prime trials against which we could compare preferred and dispreferred primes. Indeed, our hypothesis was that all of the EVA categories would prime all of the others. However, in Experiment 3 we did not observe significant between-category priming between plural trials and quantifiers/numbers. Thus the plural primes could act as a neutral control prime (a baseline) for quantifiers/numbers and vice versa. This would suggest that baseline responses for enrichment in plurals were 73%, for quantifiers 63%, and for numbers 59%. That these numbers are all above 50% confirms the intuition that the strong interpretation is the preferred sense in each case (see Grodner et al. (2010), for evidence that the strong interpretation of *some of* is the preferred sense, and reviews of the semantics of number terms for evidence that the strong interpretation of bare numerals is the preferred sense). To test whether we had an inverse preference effect we therefore compared the target following the strong prime (preferred) and the weak prime (dispreferred) to the appropriate baseline using a one sample *t*-test. For plurals, the weak prime caused mean enrichment that was significantly different from baseline, $M = 0.44$ ($SD = 0.38$), $t(93) = 7.45$, $p < .001$, but the strong prime did not, $M = 0.77$ ($SD = 0.34$), $t < 1$. The magnitude of the priming effect was significantly different across conditions, $t(93) = 3.69$, $p < .001$. Thus for plurals there was a robust inverse preference priming effect. For quantifiers, the weakly primed target was significantly different to baseline, $M = 0.41$ ($SD = 0.39$), $t(93) = 5.41$, $p < .001$, and the strong prime was marginally different to baseline, $M = 0.72$ ($SD = 0.37$), $t(93) = 1.90$, $p = .061$. Here again, the priming effect was significantly different across strong and weak conditions, $t(93) = 1.99$, $p = .050$, illustrating an inverse preference effect. Finally, for numbers, the weakly primed target differed to baseline, $M = 0.43$ ($SD = 0.39$), $t(93) = 4.14$, $p < .001$, and the strongly primed target differed marginally, $M = 0.66$ ($SD = 0.40$), $t(93) = 1.74$, $p = .084$, but the effect did not differ significantly across conditions $t(93) = 1.28$, $p = .20$. Overall then we observed the inverse preference effect for within-category priming in plural items and quantifier items, but not the number items.

Finding an inverse preference effect in our data is useful for a variety of reasons. The first is that our study uses different linguistic phenomena and somewhat different procedures to other priming studies and it is comforting that we replicate the effects of more established paradigms. It suggests that using the "better picture" option, for example, does not result in a radically different form of priming than the more standard double-picture target (e.g., Raffray & Pickering, 2010). The second is that we extend the range of priming phenomena that display the inverse preference effect. Previous work has tested inverse priming in syntactic representations (e.g., Fine & Jaeger, 2013; Hartsuiker & Kolk, 1998; Hartsuiker et al., 1990; Hartsuiker & Westenberg, 2000; Scheepers, 2003) whereas our research has demonstrated the effect in meaning-based representations. While it would be surprising if there was a fundamental difference between how syntactic and meaning-based representations were learned and used, our results suggest that meaning-based phenomena, such as EVAs, might be successfully simulated using error-based connectionist models.

Conclusion

Our aim was to better understand how people use alternatives to enrich the basic meaning of a sentence. In this respect our study makes three important contributions. (1) We demonstrate that enrichment typically thought as semantic or pragmatic can be primed in the same way syntactic properties, say, can be primed through structural priming. Models of enrichment that aim to be psychologically plausible must therefore incorporate this property into their architecture. (2) We provide data that informs theories of individual EVAs. For example, our data challenge theories of exact number semantics (e.g., Breheny, 2008; Geurts, 2006) and EVA accounts of plurals (e.g., Spector, 2007). (3) We show that there are mechanisms that are shared across different EVAs, and that some phenomena (e.g., plurals) are excluded from this class. This constrains the range of models necessary to explain how people reason with alternatives.

Appendix A

See Table A1.

Table A1
Raw cell means.

	Prime category	Target category	Prime value	Proportion strong response	S.E
Experiment 1	ADHOC	ADHOC	Strong	0.410	0.019
	ADHOC	ADHOC	Weak	0.215	0.015
	ADHOC	NUM4	Strong	0.542	0.019
	ADHOC	NUM4	Weak	0.497	0.019
	ADHOC	SOME	Strong	0.515	0.019
	ADHOC	SOME	Weak	0.492	0.019
	NUM4	ADHOC	Strong	0.319	0.018
	NUM4	ADHOC	Weak	0.283	0.017
	NUM4	NUM4	Strong	0.615	0.018
	NUM4	NUM4	Weak	0.339	0.018
	NUM4	SOME	Strong	0.553	0.019
	NUM4	SOME	Weak	0.484	0.019
	SOME	ADHOC	Strong	0.343	0.019
	SOME	ADHOC	Weak	0.301	0.017
	SOME	NUM4	Strong	0.544	0.020
	SOME	NUM4	Weak	0.474	0.019
	SOME	SOME	Strong	0.604	0.019
	SOME	SOME	Weak	0.340	0.018

Table A1 (*continued*)

	Prime category	Target category	Prime value	Proportion strong response	S.E
Experiment 2	NUM4	NUM4	Strong	0.755	0.022
	NUM4	NUM4	Weak	0.549	0.027
	NUM4	NUM6	Strong	0.707	0.024
	NUM4	NUM6	Weak	0.587	0.026
	NUM4	SOME	Strong	0.664	0.025
	NUM4	SOME	Weak	0.642	0.026
	NUM6	NUM4	Strong	0.698	0.024
	NUM6	NUM4	Weak	0.624	0.026
	NUM6	NUM6	Strong	0.709	0.024
	NUM6	NUM6	Weak	0.550	0.027
	NUM6	SOME	Strong	0.667	0.025
	NUM6	SOME	Weak	0.667	0.026
	SOME	NUM4	Strong	0.710	0.024
	SOME	NUM4	Weak	0.668	0.025
	SOME	NUM6	Strong	0.675	0.025
	SOME	NUM6	Weak	0.652	0.025
	SOME	SOME	Strong	0.713	0.024
	SOME	SOME	Weak	0.458	0.027
Experiment 3	NUM4	NUM4	Strong	0.668	0.025
	NUM4	NUM4	Weak	0.422	0.026
	NUM4	PLURALS	Strong	0.739	0.023
	NUM4	PLURALS	Weak	0.733	0.024
	NUM4	SOME	Strong	0.687	0.024
	NUM4	SOME	Weak	0.631	0.026
	PLURALS	NUM4	Strong	0.594	0.025
	PLURALS	NUM4	Weak	0.587	0.026
	PLURALS	PLURALS	Strong	0.774	0.022
	PLURALS	PLURALS	Weak	0.439	0.026
	PLURALS	SOME	Strong	0.636	0.025
	PLURALS	SOME	Weak	0.628	0.025
	SOME	NUM4	Strong	0.642	0.026
	SOME	NUM4	Weak	0.568	0.026
	SOME	PLURALS	Strong	0.738	0.024
	SOME	PLURALS	Weak	0.712	0.024
	SOME	SOME	Strong	0.730	0.024
	SOME	SOME	Weak	0.413	0.026

References

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*, 247–264.

Baayen, R. H. (2011). LanguageR: Data sets and functions with "Analyzing linguistic data: A practical introduction to statistics". *R Package Version, 1*, 4.

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition, 118*, 84–93.

Bates, D., Maechler, M., & Bolker, B. (2011). lme4: Linear mixed-effects models using S4 classes.

Bergen, L., & Grodner, D. J. (2012). Speaker knowledge influences the comprehension of pragmatic inferences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 1450.

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18*, 355–387.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language, 51*, 437–457.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition, 75*, B13–B25.

Breheny, R. (2008). A new look at the semantics and pragmatics of numerically quantified noun phrases. *Journal of Semantics, 25*, 93–140.

Breheny, R., Katsos, N., & Williams, J. (2006). Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. *Cognition, 100*, 434–463.

Carston, R. (1998). Informativeness, relevance and scalar implicature. *Pragmatics and Beyond*, 179–238.

Chang, F., Dell, G. S., & Bock, J. K. (2006). Becoming syntactic. *Psychological Review, 113*(2), 234–272.

Chierchia, G. (2004). Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In A. Belletti (Ed.), *Structures and beyond*.

Chierchia, G., Fox, D., & Spector, B. (2012). Scalar implicature as a grammatical phenomenon. In P. Portner, C. Maienborn, & K. von Heusinger (Eds.). *An international handbook of natural language meaning* (Vol. 3, pp. 2297–2331). Mouton de Gruyter.

Cleland, A. A., & Pickering, M. J. (2003). The use of lexical and syntactic information in language production: Evidence from the priming of noun-phrase structure. *Journal of Memory and Language, 49*(2), 214–230.

Crain, S., & Thornton, R. (1998). *Investigations in universal grammar*. Cambridge, MA: MIT Press.

Degen, J., & Tanenhaus, M. K. (2015). Processing scalar implicature: A constraint-based approach. *Cognitive Science, 39*, 667–710.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*, 1117–1121.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6*, 274–290.

Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 1–17. http://dx.doi.org/10.3389/fpsyg.2014.00781.

Elman, J. L., Hare, M., & McRae, K. (2004). Cues, constraints, and competition in sentence processing. In M. Tomasello & D. Slobin (Eds.), *Beyond nature–nature: Essays in honor of Elizabeth Bates* (pp. 111–138). Mahwah, NJ: Lawernce Erlbaum Associates.

Fine, A. B., & Jaeger, T. F. (2013). Evidence for implicit learning in syntactic comprehension. *Cognitive Science*, 1–14.

Fox, D. (2006). Free choice disjunction and the theory of scalar implicatures. In Uli. Sauerland & Penka. Stateva (Eds.), *Presupposition and implicature in compositional semantics* (pp. 71–120). New York: Palgrave Macmillan.

Franke, M. (2011). Quantity implicatures, exhaustive interpretation, and rational conversation. *Semantics and Pragmatics, 4*, 1–82.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition, and logical form*. New York: Academic Press.

Geurts, B. (2006). Take "five": The meaning and use of a number word. In V. Svetlana & L. Tasmowski (Eds.), *Non-definiteness and plurality* (pp. 311–329). Amsterdam: Benjamins.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.

Grodner, D., Klein, N. M., Carbary, K. M., & Tanenhaus, M. K. (2010). "Some", and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition, 116*, 42–55.

Grodner, D., & Sedivy, J. (2011). The effect of speaker-specific information on pragmatic inferences. *The processing and acquisition of reference, 2327*, 239–272.

Guasti, T. M., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and cognitive processes, 20*(5), 667–696.

Hartsuiker, R. J., & Kolk, H. H. (1998). Syntactic persistence in Dutch. *Language and Speech, 41*, 143–184.

Hartsuiker, R. J., Kolk, H. H., & Huiskamp, P. (1990). Priming word order in sentence production. *Quarterly Journal of Experimental Psychology, 52A*, 129–147.

Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition, 75*, B27–B39.

Hirschberg, J. (1991). *A theory of scalar implicature*.New York: Garland Press.

Horn, L. R. (1972). *On the semantic properties of logical operators in English*. Ph.D. thesis, University of California, Los Angeles.

Horn, L. R. (1989). *A natural history of negation*.Chicago, Ill.: University of Chicago Press.

Horn, L. (1992). The said and the unsaid. *Ohio State University Working Papers in Linguistics, 40*, 163–192.

Huang, Y. T., & Snedeker, J. (2009). On-line interpretation of scalar quantifiers: Insight into the semantic-pragmatics interface. *Cognitive Psychology, 58*, 376–415.

Huang, Y. T., Spelke, E., & Snedeker, J. (2013). What exactly do numbers mean? *Language Learning and Development, 9*, 105–129.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59*, 434–446.

Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition, 127*, 57–83.

Johnson-Laird, P. N., & Bara, B. G. (1984). Syllogistic inference. *Cognition, 16*, 1–61.

Katsos, N., & Bishop, D. V. (2011). Pragmatic tolerance: Implications for the acquisition of informativeness and implicature. *Cognition, 120*(1), 67–81.

Katzir, R. (2007). Structurally-defined alternatives. *Linguistics and Philosophy, 30*, 669–690.

Kennedy, C. (2013). A scalar semantics for scalar readings of number words. In Ivano Caponigro & Carlo Cecchetto (Eds.), *From grammar to meaning: The spontaneous logicality of language* (pp. 172–200). Cambridge University Press.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

Levinson, S. C. (1983). *Pragmatics. Cambridge textbooks in linguistics*. Cambridge University Press.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., . . . Wagenmakers, E.-J. (2015). JASP (Version 0.7) [Computer software].

Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua, 133*, 152–163.

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. (1998). Modeling thematic fit (and other constraints) within an integration competition framework. *Journal of Memory and Language, 38*, 283–312.

Musolino, J. (2004). The semantics and acquisition of number words: Integrating linguistic and developmental perspectives. *Cognition, 93* (1), 1–41.

Newstead, S. E., Pollard, P. E., & Allen, J. L. (1992). The source of belief bias effects in syllogistic reasoning. *Cognition, 45*, 257–284.

Papafragou, A., & Musolino, J. (2003). Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition, 86*(3), 253–282.

Patson, N. D., & Ferreira, F. (2009). Conceptual plural information is used to guide early parsing decisions: Evidence from garden-path sentences with reciprocal verbs. *Journal of Memory and Language, 60*, 464–486.

Patson, N. D., George, G. E., & Warren, T. (2014). The conceptual representation of number. *Quarterly Journal of Experimental Psychology, 67*, 1349–1365.

Patson, N. D., & Warren, T. (2011). Building complex reference objects from dual sets. *Journal of memory and language, 64*(4), 443–459.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language, 39*, 633–651.

Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: A critical review. *Psychological Bulletin, 134*, 427.

R Core Team (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing <http://www.R-project.org/>.

Raffray, C. N., & Pickering, M. J. (2010). How do people construct logical form during language comprehension? *Psychological Science, 21*, 1090–1197.

Raffray, C. N., Pickering, M. J., Cai, Z. G., & Branigan, H. P. (2014). The production of coerced expressions: Evidence from priming. *Journal of Memory and Language, 74*, 91–106.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237.

Sauerland, U. (2003). A new semantics for number. *Semantics and Linguistic Theory*, 258–275.

Sauerland, U. (2004). Scalar implicatures in complex sentences. *Linguistics and Philosophy, 27*, 367–391.

Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition, 89*, 179–205.

Spector, B. (2007). Aspects of the pragmatics of plural morphology: On higher-order implicatures. In U. Sauerland & P. Stateva (Eds.), *Presuppositions and implicatures in compositional semantics*. Palgrave-Macmillan.

Spector, B. (2013). Bare numerals and scalar implicatures. *Language and Linguistics Compass, 7*, 273–294.

Thothathiri, M., & Snedeker, J. (2008). Give and take: Syntactic priming during spoken language comprehension. *Cognition, 108*, 51–68.

Tieu, L., Romoli, J., Zhou, P., & Crain, S. (in press). Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics*, doi:http://dx.doi.org/10.1093/jos/ffv001.

Tomlinson, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of Memory and Language, 69*, 18–35.

van Rooij, R., & Schulz, K. (2004). Exhaustive interpretation of complex sentences. *Journal of Logic, Language and Information, 13*, 491–519.

van Tiel, B., van Miltenburg, E., Zevakhina, N., & Geurts, B. (2014). Scalar diversity. *Journal of Semantics*. http://dx.doi.org/10.1093/jos/ffu017.