COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 39 (2015) 1646–1674

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12212

Hey Little Sister, Who's *the Only One*? Modulating Informativeness in the Resolution of Privative Ambiguity

Francesca Foppolo,[a] Marco Marelli,[b] Luisa Meroni,[c] Andrea Gualmini[d]

[a]*Psychology Department, University of Milano-Bicocca*
[b]*Center for Mind/Brain Sciences, University of Trento*
[c]*Utrecht Institute of Linguistics, OTS (Utrecht University)*
[d]*Damp en zo BV, Utrecht*

Received 10 June 2013; received in revised form 8 July 2014; accepted 15 September 2014

Abstract

We present two eye-tracking experiments on the interpretation of sentences like "The tall girl is (not) *the only one* that . . .," which are ambiguous between the anaphoric (the only girl that . . .) and the exophoric interpretation (the only individual that . . .). These interpretations differ in informativeness: in a positive context, the exophoric (strong) reading entails the anaphoric (weak), while in a negative context the entailment pattern is reversed and the anaphoric reading is the strongest one. We tested whether adults rely on considerations about informativeness in solving the ambiguity. The results show that participants interpreted *one* exophorically in both positive and negative contexts. Given these findings, we cast doubts on the idea that Informativeness plays a role in ambiguity resolution and proposes a Principle of Maximal Exploitation: When a context is provided, adults extend their domain of evaluation to include the whole scenario, independently from truth-conditional considerations about informativity and strength.

Keywords: Visual world paradigm; Semantic ambiguity; Anaphora; Only; Entailment

1. Introduction

In this paper, we intend to address a central question in the processing of semantic ambiguity by means of two eye-tracking studies. As a test case, we use the phenomenon of privative ambiguities, that is, ambiguous sentences where one reading truth-conditionally entails the other (Horn, 1996). To exemplify, consider (1), uttered, for example, in front of a playground in which some boys and some girls are playing together:

Correspondence should be sent to Francesca Foppolo, Department of Psychology, University of Milano-Bicocca, P.zza dell'Ateneo Nuovo 1, 20126 Milan, Italy. E-mail: francesca.foppolo@unimib.it

(1) The tall girl is *the only one* that is wearing a hat

Sentence (1) contains an ambiguity, ensuing from the fact that the expression *the only one* might be interpreted *anaphorically*, that is, as referring to the linguistic antecedent in the head noun of the subject NP as in (2a), or *exophorically*, thereby making reference to the wider discourse domain, as in (2b). Crucially, these two interpretations stand in a subset/superset relation, as in (2c):

(2) (a) The tall girl is the only *girl* that is wearing a hat *anaphoric*
(b) The tall girl is the only *child* that is wearing a hat *exophoric*
(c) scale of informativeness: (a) < (b); entailing pattern: (b) entails (a)

The information in (2c) tells us that, whenever (1) is true under interpretation (2b), it is also true under interpretation (2a), but not vice versa: If the tall girl is the only *individual* in the domain of discourse that is wearing a hat, it necessarily follows that she will also be the only *girl* that is wearing a hat. Conversely, if the tall girl is the only *girl* that is wearing a hat, there might be another individual in the domain of discourse, a boy for instance, that is wearing a hat, too. This, however, would not make sentence (1) false under the anaphoric reading (2a). In terms of *informativeness*, interpretation (2b) of an affirmative sentence like (1) is the *strong* interpretation, provided that it requires more commitments for its truth: In this case, for the sentence to be true under this reading, all the individuals in the discourse domain must be checked, and any other child (boy or girl) with a hat, except for the individual mentioned in the sentence, would make the sentence false under the exophoric reading. Conversely, fewer commitments are required for the sentence to be true under the anaphoric reading (2a); in this case, only a subset of the individuals in the discourse domain must be checked (in the example, only the girls within the group of children). For affirmative sentences, this is the *weak* interpretation, provided that it makes the sentence true in a wider set of circumstances: Under the anaphoric reading, not only the sentence would be true in all cases in which no other girl in the domain of discourse except for the one alluded to in the sentence is wearing a hat; it would also be true in a situation in which one of the boys is wearing a hat, too.

Many linguistic phenomena exhibit this property and have been extensively debated within theories dealing with the semantic/pragmatic interface. For example, the Strongest Meaning Hypothesis has been advocated to account for the interpretation of reciprocals or plural predicates (Dalrymple, Kanazawa, Kim, Mchombo, & Peters, 1998; Winter, 2001) and the Principle of Pragmatic Strengthening was originally proposed by Krifka (1996) to account for preferences in the resolution of plural predication and the so-called donkey sentences. These principles assume that, whenever alternative readings are allowed by the grammar, one should favor the strong over the weak one (if consistent with general background assumptions). Consider, for example, a case of plural predication: "The boys are running and waving." Under the strong interpretation, which allegedly is the preferred one according to the Principle of Pragmatic Strengthening, the sentence would describe a situation in which all the boys are running and all the boys are waving,

but not a situation in which some of the boys are running and some others are waving. This would be the dispreferred weak reading of this sentence. A similar argument was also implemented in Chierchia (2006; see also Chierchia, Fox, & Spector, 2013) to account for the interpretation of sentences involving scalar terms like *some* and polarity-sensitive items like *ever* and *any*, whose interpretation involves alternatives and considerations about informativeness. For example, the sentence "The girl ate *some* of the cookies" is normally assigned its strong interpretation: In upward entailing contexts, this coincides with the pragmatic enriched meaning "The girl ate *some but not all* of the cookies." This is more informative than "The girl ate *some (and possibly all*) of the cookies," that would be true if the girl ate all of the cookies. In downward entailing contexts like the antecedent of conditionals (in which the direction of entailment between alternatives is reversed), the scalar inference is suspended in favor of the more informative meaning "If the girl ate *some or all* of the cookies she will gain weight."[1] For the sake of simplicity, we will subsume these principles under the general label of Principle of Maximizing Informativity, which predicts that adults will always favor the strong interpretation of a sentence containing a privative ambiguity (i.e., the one that is true in narrower sets of circumstances), so as to convey the most informative message.

However, another widespread assumption goes in the opposite direction: In the philosophical literature, the Principle of Charity (Davidson, 1984; Quine, 1964) is a methodological principle which "constraints the interpreter to maximize the truth or rationality in the subject's saying" (Blackburn, 1994: 62). This principle is especially invoked to make sense of sentences whose meanings are not uniquely determined, as in the case of ambiguous sentences. In these cases, the principle would lead subjects to view a sentence as true as long as one of its possible interpretations is true. Similar principles have also been discussed by semanticists working on scope ambiguities (e.g., Abusch, 1993; Meyer & Sauerland, 2009; Reinhart, 1997).

Despite the fact that they make opposite predictions, both the principles mentioned above refer to the general notion of informativity and strength according to which, whenever alternatives might be ordered with respect to the amount of information they provide, a preference for the strong or the weak reading is expected, independently of context. Following Geurts (2000), we will dub these accounts under a general Informativeness Principle, abstracting from the polarity of the predictions they make regarding speakers' preferences.

As we have seen, the idea that entailment relations between alternative readings play a role in ambiguity resolution provided solid grounds for semantic/pragmatic theorizing. At the same time, though, this idea has mostly been defended on purely logical grounds on the basis of introspective data,[2] with all the limitations that this method implies (cf. Geurts, 2009 and Geurts & Pouscoulous, 2009 for a thorough discussion about introspection as a tool of gathering data on pragmatic inferences). Notable exceptions are Yoon's experimental work on plural predicates (1996), Geurts' work on so-called donkey sentences (Geurts, 2002), and a study by Crain, Ni, and Conway (1994) that specifically addressed the question of the resolution of privative ambiguity related to the phenomenon of one-substitution, and that constituted the starting point of our research. In this latter

study, Crain and colleagues tested sentences like "The big elephant is *the only one* playing guitar" with 3- to 5-year-old English-speaking children by means of a Truth-Value-Judgement task (Crain & McKee, 1985). They asked children to evaluate this sentence in a scenario in which, in addition to the big elephant, that was the only elephant playing the guitar, an octopus and a bird were playing a guitar, too. This made the sentence true under the anaphoric interpretation ("The big elephant is the only *elephant* playing guitar") but false under the exophoric interpretation ("The big elephant is the only *animal in the domain* playing guitar"). What they found is that children consistently rejected the target sentence on the ground that other animals were performing the same action as the big elephant. Crain et al. also tested a small group of adults on the same sentences but using a different task: Adults were asked to imagine a scenario that made those sentences false. What they found is that adults tended to mention another animal belonging to the same individual-level category of the antecedent already mentioned in the sentence. In the example provided above, adults imagined a situation in which there was another elephant that was playing a guitar too instead of mentioning an animal of a different species, thus interpreting *the only one* anaphorically. On the basis of these results, Crain and colleagues conjectured that, in dealing with semantic ambiguity, the adult parser conforms to what they named a Minimal Commitment strategy. This term refers to adults' alleged preference for the weak alternative in solving privative ambiguity, a preference that is ultimately motivated by the need to avoid unnecessary commitments that might need to be revised at a later stage. In a recent contribution, we argued against this conclusion and hypothesized that the difference observed in the strategy adopted by children and adults in solving the ambiguity related to *the only one* might simply be due to the difference in the tasks that the two populations were administered (Foppolo, Meroni, Marelli, & Gualmini, 2014). In particular, we hypothesized that the strategy of Minimal Commitment observed in adults might depend on the fact that no context was provided for falsification in the Crain et al.'s study. By adding a context before the target sentences, we showed that the probability of mentioning an individual belonging to a different category than the linguistic antecedent was significantly higher when other entities were mentioned in the preceding context.

In this paper, we pursued this idea further by means of two eye-tracking studies exploiting the visual world paradigm (Cooper, 1974) in which participants' eye movements were recorded during on-line sentence processing while exploring a visual scene. This technique has been largely used in the past years for the investigation of a variety of psycholinguistic phenomena. Since pioneer works by Altmann and Kamide (1999) and Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995), the recording of eye movements has shown that listeners can anticipate toward an intended target before it is actually mentioned in the sentence and that they actually solve syntactic ambiguities while the sentence unfolds. Different studies reported evidence that the on-line interpretation of linguistic input might be affected by diverse informational sources, either linguistics or contextual. For example, the presence or absence of a contrast object in the visual scene has been shown to affect the resolution of syntactic and semantic ambiguity (Sedivy, Tanenhaus, Chambers, & Carlson, 1999; Tanenhaus et al., 1995); and properties of the

verb have been shown to influence the resolution of structural ambiguity (Kamide, Scheepers, & Altmann, 2003).

Following Crain et al.'s study, we tested the ambiguity associated with the expression *the only one*, exemplified in (1). Our aim was to disentangle the contribution of the different strategies or processing principles proposed in the literature to explain how adults solve this ambiguity. We tested ambiguous sentences like "The big X is *the only one* that is doing P" in a scenario that comprised four different subsets of abstract objects, for example, four sets of different shapes of four objects each (e.g., four squares, four circles, four pentagons, and four triangles in the same scenario) in which only one X (but some other Y) was doing P. The sentence could be judged true if the anaphoric interpretation was assigned to *one* (in fact, the big X was always the only X that was doing P) and false if it was assigned the exophoric interpretation (in fact, the big X was not the only individual that was doing P, provided that some Y was doing P as well). We first considered the truth-value judgment assigned to these ambiguous sentences containing a privative ambiguity that is true under the weak reading and false under the strong reading, by virtue of the entailing relation that holds between their alternative readings (cf. [2]). Also, we evaluated two measures of processing: First of all, we measured the response time in providing a truth-value judgment on these sentences; secondly, we recorded the eye movements employed in the evaluation process, which arguably indicate the domain restriction adopted to interpret the pronoun: either the entire set of individuals in the domain if the exophoric interpretation is adopted (e.g., the entire set of children in example [1]), or the subset of individuals belonging to the same individual-level category of the individual mentioned in the sentence if the anaphoric interpretation is adopted (e.g., only the girls among all children). In this respect, the exploration pattern during sentence evaluation is informative of participants' parsing preferences independently from the final judgement assigned to the ambiguous sentence, which might reflect post-grammatical processes. Indeed, listeners might provide a final judgment that results from a decision made a posteriori, and possibly after one of the alternative readings had been considered but rejected. In this case, their eye movements would reveal the paths that lay behind the whole decisional process, showing which alternatives have been considered (if any) and, eventually, in which order. In our study, in particular, we evaluated how far listeners went in search for a referent for *one* and how thoroughly they explored the scenario to restrict the domain of interpretation of the anaphoric expression before providing a truth-value judgment.

Unlike other studies employing the visual world paradigm, in our study we kept the contextual information constant and modulated the polarity of the sentence to test whether and how informativeness considerations about alternative readings of the ambiguity affect the exploration of the scenario and, eventually, which interpretation is considered first. To investigate the type of exploration adopted during the on-line processing of our linguistic stimuli, we analyzed the eye movements in the critical ambiguous sentences comparing them with two types of unambiguous control trials that required a global or a local exploration of the scenario in order to be evaluated, paralleling the exophoric/anaphoric contrast in the domain restriction in the interpretation of the pronoun. In one case, control trials required the exploration of the entire scenario, matching the

exophoric interpretation of the ambiguous sentences in which all the individuals in the domain need to be checked to see if, among all, the X mentioned in the sentence was the only individual that is doing P. The other type of control trials required the exploration of a specific subdomain of individuals (those of the same individual-level category of the individual mentioned in the sentence). These matched the anaphoric interpretation of *one*, in which only the individuals of the same individual-level category of the linguistic antecedent needed to be checked to evaluate if, among them, X was the only X doing P, independently from the fact that other individuals of a different category in the domain of discourse were doing P as well.

To test whether the process of exploration and evaluation proceeded on the basis of truth-value considerations, as maintained by the Informativeness principle, we modulated the context in which the anaphora appeared across the two experiments, so as to make the anaphoric interpretation true in Experiment 1 and false in Experiment 2. This was achieved by inserting negation, so as to reverse the entailing relation between alternative interpretations and make the anaphoric reading the strong one (as in the previous scenario, it is true that the big X is *not* the only individual that is doing P, provided that one Y is doing P as well; however, it is *not* true that the big X is *not* the only X that is doing P, because in fact no other X is doing P).

The idea at the basis of this manipulation is very simple: (a) if participants relied on the Principle of Maximizing Informativity, they would shift from the exophoric to the anaphoric interpretation of the anaphora across experiments, provided that the exophoric interpretation is the strong one in the declarative affirmative sentences used in Experiment 1 while the anaphoric interpretation is the strong one in the declarative negative sentences used in Experiment 2; (b) if participants were guided by the Principle of Charity, they would always opt for the interpretation that renders the sentence true. This means that we would predict the inverse path across experiments: Following the Principle of Charity, listeners would shift from the anaphoric to the exophoric interpretation, provided that the anaphoric interpretation is the weaker one in the declarative affirmative sentences used in Experiment 1 and the exophoric interpretation is the weaker one in the declarative negative sentences used in Experiment 2. In both these cases, a shift in interpretation is predicted between Experiments 1 and 2. However, a third possibility exists, that people do not rely (solely) on informativeness to solve (privative) ambiguity but consider other elements in the evaluation of the alternatives.

To solve this issue, our experimental question is the following: What do people rely on in solving ambiguity, if informativeness is not at stake?

2. Experiment 1: Do adults minimally commit themselves and opt for the weak alternative?

In the first experiment, we investigated how adults interpreted and judged Italian sentences like (3), ambiguous between interpretations (4a) and (4b):

(3) Il quadrato piccolo è l'unico che fa il poliziotto
the square small is the only-one that makes the policeman
The small square is the only one that is a policeman

(4) (a) The small square is the only *square* that is a policeman *anaphoric*
(b) The small square is the only *thing/shape* that is a policeman *exophoric*

A visual scenario was provided for evaluation in which, crucially, the anaphoric (weak) interpretation was true but the exophoric (strong) interpretation was false. The task employed was a picture Truth-Value Judgment task similar to that used by Crain et al. with children, with the addition that, in this case, we also recorded eye movements and reaction time data. Our main goal was to gather new insights on the way adults solved semantic ambiguity when they are provided with a context for evaluation. In particular, we aimed at testing whether adults selected the weak alternative in case of ambiguity, following a general Principle of Charity and replicating the results obtained by Crain et al. in the Sentence Falsification task, or whether they favored the strongest among alternatives, following a general Principle of Maximizing Informativity. Provided that the scenario given in Experiment 1 was only compatible with the weak (anaphoric) interpretation of the ambiguous expression, the Principle of Charity predicts that adults judge the sentences true, solving the ambiguity anaphorically as in (4a), while the Principle of Maximizing Informativity predicts that they judge the sentences false, solving the ambiguity exophorically as in (4b).

As far as the exploration pattern is concerned, the Principle of Charity predicts that participants focus only on the objects of the same individual-level category of the linguistic antecedent (i.e., the squares) when processing sentence (3), independently of what happens to non-square individuals. Conversely, the Principle of Maximizing Informativity predicts that participants consider the whole domain of individuals presented on the screen, thereby taking into consideration other non-square individuals that are performing the same action as the subject NP of the target sentence.

2.1. Methods

2.1.1. Participants

Thirty-one students of the University of Milano-Bicocca participated in the experiment for course credits. One subject was excluded because of technical problems in the eye-tracking recording, so the analyses only include 30 subjects. All participants were native speakers of Italian, had normal or corrected-to-normal vision, and were naive with respect to the goals of the experiment.

2.1.2. Materials

We used a Visual World Paradigm (Cooper, 1974; Huettig, Rommers, & Meyer, 2011; Tanenhaus et al., 1995), in which subjects were asked to judge a sentence as true or false with respect to a static visual scene while their eye movements were recorded. Materials comprised 82 trials, including 30 critical trials, 39 control trials, four warm-ups, and nine

fillers (a complete list of materials is given in the Appendix S1). Each trial consisted of an Italian sentence, auditorily presented, and a visual scenario. The order of presentation of the experimental trials was fully randomized across participants.

Critical trials were ambiguous sentences like (3) in which a noun (e.g. *square*) was combined with an adjective (e.g. *small*) and associated with a predicate (e.g. *is a policeman*). The noun was the name of a geometrical figure (*square*, *triangle*, *pentagon*, *circle*), a letter or a digit; the adjectives was either a scalar adjective (*big* or *small*) or a color adjective (*green*, *blue*, *yellow*, *red*); the predicate described an action (playing tennis, reading, etc.), a mental/emotional status (being happy, thinking, etc.), or a state of affair (being a policeman, wearing a bow tie, being a cook, etc.). Predicates were represented visually by an emoticon, that is, a pictorial representation of a facial expression commonly used in instant messaging. All critical sentences contained the ambiguous expression *l'unico* or *l'unica* (*the only-one*$_{[+MASC/+FEM]}$). The ambiguous expression *unico/a* had to be marked for gender (masculine or feminine), depending on what kind of items were presented on the screen: It is always feminine in case of letters and always masculine in case of shapes and numbers. The presence of gender marking, however, could not be used to solve the ambiguity associated with the expression *the only one*, given that all the items in the same scenario were of the same gender. The crucial factor was that, independently of gender marking, the critical sentences could be interpreted anaphorically (i.e., by making reference to the linguistic antecedent) or exophorically (i.e., by making reference to a wider discourse domain). This, in turn, had an effect on sentence truth-values: All sentences in Experiment 1 were in fact true under the anaphoric reading and false under the exophoric one. To clarify, let us consider sentence (3), which is repeated in (5) for convenience in its English version, accompanied by its alternative interpretations.

(5) The small square is *the only one* that is a policeman
  (a) The small square is the only *square* that is a policeman *anaphoric*
  (b) The small square is not the only *thing/shape* that is a policeman *exophoric*
  (c) scale of informativeness: (a) < (b); entailing pattern: (b) entails (a)

Sentence (5) had to be evaluated in the scenario represented in Fig. 1.

Crucially, sentence (5) was true in the scenario only if the ambiguous segment *the only one* was interpreted anaphorically, that is, by making reference to the linguistic antecedent, as in (5a). If interpreted exophorically as in (5b), the same sentence would be false in the same scenario, given that a circle and a triangle are policemen, too (cf. Fig. 1). Please note that the individual-level category of the objects within each quadrant did not vary, but it varied across quadrants, though remaining within the restriction of the superorder category. For example, the 16 objects in the scenario in Fig. 1 were all shapes, but of four different types (one for each of the four quadrants) and all comprised four objects each: four pentagons in the top-left quadrant, four circles in the bottom-left quadrant, four triangles in the top-right quadrant, and four squares in the bottom-right quadrant. This is crucial to understand the pattern of exploration of the scenario revealed by eye movements.

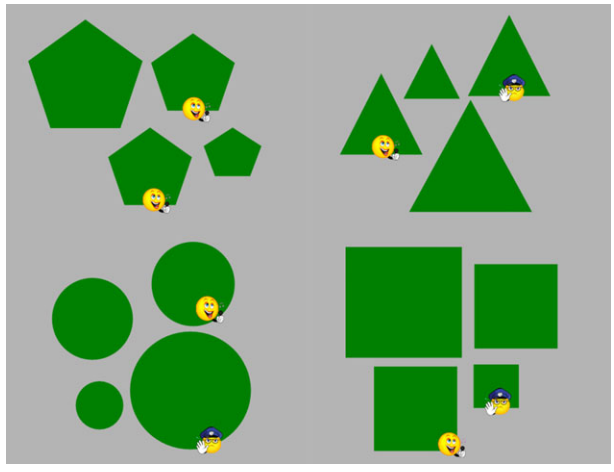

Fig. 1. Visual scenario for the sentence: "The small square is the only one that is a policeman."

Thirty-nine control trials were included (18 false and 21 true) and varied with respect to the type of exploration required to judge them, thus paralleling the anaphoric/exophoric distinction in the critical trials. In 20 control trials, a target quadrant could be identified and the exploration of this single quadrant sufficed to decide whether the sentence was true or false. These control trials were labeled One-Quadrant (ONEq), 11 of which were true. The remaining 19 control trials were labeled ALL-Quadrants (ALLq), 10 of which were true; in these, a target quadrant could not be identified and thus the exploration of the scene could not be restricted to a specific quadrant in order to provide a truth-value judgment, but required the exploration of the entire scene. To exemplify, ONEq control trials were sentences like (6), and ALLq control trials were sentences like (7), uttered in the scenarios in Fig. 2 (in which both sentences are true):

(6) Tutti i cinque sono cuochi
All the fives are cooks

(7) Nessuna figura grande indossa un farfallino
no shape big wears a bow tie
No big shape wears a bow tie

The peculiarity of ONEq control trials like (6) is that, in order to evaluate them, listeners need to explore thoroughly only the single quadrant that contains the fives, once this is individuated (in this case, the top-left quadrant in Fig. 2, left). This quadrant can be identified once the subject of the sentence is heard (in this case, "All the fives") and no other quadrant has to be checked to decide about the truth/falsity of the sentence (in this case, the truth). In case of (7), instead, no target quadrant can be identified after hearing the phrasal subject (in this case, "No big shape"), given that every quadrant contains one shape that is bigger than the others within that quadrant. For this reason, a truth-value

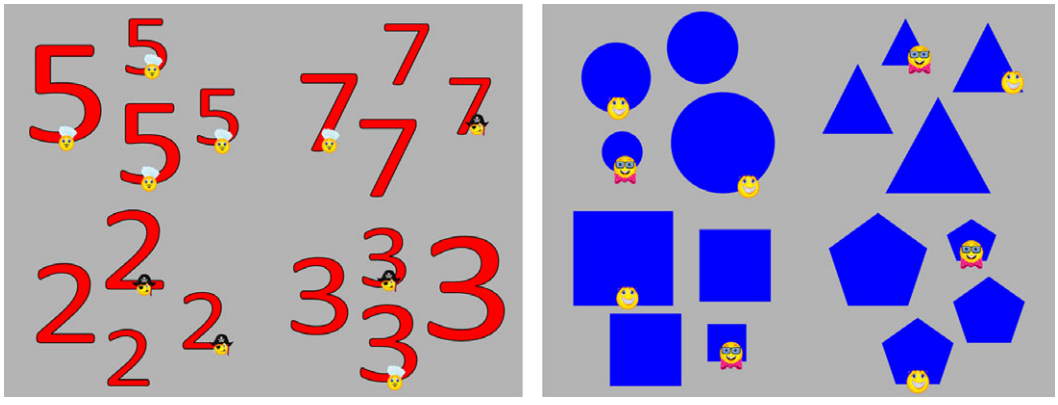

Fig. 2. On the left: visual scenario for sentence (6) "All the fives are cooks" (ONEq control, true). On the right: visual scenario for sentence (7) "No big shape wears a bow tie" (ALLq control, true).

judgment (in this case, true) can be provided only after the whole scenario is explored thoroughly and, in this case, the biggest shape in each quadrant is checked (Fig. 2, right).

Additional nine fillers were interspersed among critical and control trials. These filler trials were unrelated to the experimental question and for this reason they were not entered in the statistical analysis.

2.1.3. Predictions

Critical sentences were ambiguous between the anaphoric/exophoric interpretation of the expression *the only one*. As we said, these interpretation are not truth-conditionally independent from one another (cf. [5]), being the exophoric interpretation the strong one in case of Experiment 1. Assuming that participants rely on a general Informativeness Principle, a clear-cut prediction can be thus made with respect to truth-value judgments: (a) if participants follow the Principle of Maximizing Informativity, the answer FALSE is expected and the exophoric interpretation adopted, and (b) if participants follow the Principle of Charity, the answer TRUE is expected and the anaphoric interpretation adopted. As a consequence, we predicted the following pattern in participants' eye movements: (a) in the former case, if the exophoric interpretation is adopted, the critical ambiguous sentences (AMB) should pattern differently from ONEq control trials, provided that the whole scenario would be checked before a decision is made, and (b) in the latter case, if the anaphoric interpretation is adopted, the critical ambiguous sentences (AMB) should pattern more like ONEq control trials, as listeners should focus on one single quadrant and not consider the remaining quadrants once the target has been identified. For the sake of clarity, we summarize these predictions in Table 1.

2.1.4. Technical details

An EyeLink 1000 eye-tracker manufactured by SR Research Ltd (Ottawa, Ontario, Canada). was used to monitor the participants' eye movements. Subjects were seated at a desk with their eyes at about 60 cm from the display and with their head leaned against a headrest, while a desktop camera sampled the pupil position and size at a frequency of

Table 1
Predictions of the two principles tested in Experiment 1 for truth-value judgments, interpretation, and exploration pattern

	Maximizing Informativity	Charity
Truth-value judgment	FALSE	TRUE
Interpretation of *one*	exophoric	anaphoric
Exploration pattern	AMB $\neq$ ONEq	AMB = ONEq

500 Hz. The recording was monocular. Stimuli were presented on a PC running the Experiment Builder software package (Longhurst, 2006). The 13 spherical random option was used for calibrating the eye-tracker. A drift control was done before each trial and re-calibration was performed when necessary. Participants provided their judgements by pressing a button for true and a button for false.

The screen was virtually divided into four identical quadrants for analysis purposes, but no explicit marking of separation among the quadrants was visible to participants. In each quadrant, four different objects were shown, belonging to the same individual-level category, differing either in size or color; as we said, all the 16 objects on the screen (four in each quadrant) belonged to the same super-order category (shapes, letters, or digits). In case objects varied in size within each quadrant, their color was kept constant across quadrants and varied randomly within trials among blue, yellow, green, and red; in case objects varied in color within the single quadrant, their size was kept constant. Separate 472 $\times$ 359 images were created for each quadrant, each consisting of four different objects supplied with emoticons on an 180,180,180 RGB gray rectangular background. The final location of each image on the screen was pseudo-randomized across trials. Of the 30 critical trials, 15 varied in size as in Fig. 1 (and were described by a sentence containing a scalar adjective, like [5]) and 15 varied in color and were described by a sentence containing a color adjective. Roughly the same ratio between scalar and color variation was maintained for control and filler trials. As we said, the objects were associated with emoticons to represent predicates. The total number of emoticons in each screen was always 8 in case of critical trials and varied between 8 and 15 in case of control trials. Only two different predicates were represented in each trial: the predicate mentioned in the sentence and a competitor.

Sentences were presented auditorily through headphones. They were recorded by a female native Italian with the most neutral prosody possible so as to insure that no portion of the sentence received contrastive or any other form of marked stress. Peaks were normalized using Audacity. Sentence onset was time-locked to the appearance of the visual stimuli on the screen.

2.1.5. Procedure

Participants were first informed about how the eye-tracker and the response box worked and then they were given instructions about the task. In particular, they were told to listen to the sentences carefully while looking at the screen and then to decide whether these sentences were true or false relatively to the scenario they were looking at. In doing

so, they were asked to be as accurate as possible in their judgments, but, at the same time, they were prompted to be fast and intuitive. To prepare participants to the unexpected situation in which abstract objects like letters, digits, and shapes were "doing things," they were explicitly informed about the structure of our stimuli and were asked to imagine that each scenario represented a snapshot of an alien world in which something different was going on each time. To familiarize them with the procedure and the materials, we provided participants with two practice trials, one true and one false, in which we explicitly trained them to consider the emoticons as denoting the actions that these unusual characters were performing in that moment, in that world.[3]

The experiment was run in a single session that lasted about 15–20 min. Participants had the opportunity to rest their eyes before each trial during the drift-control procedure, by closing them or simply looking around a bit. They were also informed that they could come off the headrest anytime during the experiment, but none of the subjects felt the need to do so. Each trial was automatically terminated when the participant pressed the response button, and a gray screen with a black circle in the center was shown for drift-correction before the next trial.

2.2. Results

We analyzed fillers and control trials to check for accuracy first. Overall, participants responded correctly 96% of the time (*SD* 4%). No subject was excluded on the basis of bad performance (the minimum participants' score was 89% correct). We excluded responses on fillers and all the incorrect answers on control trials (<4%) from further analyses. Then, we examined the truth-value judgments provided by our participants in case of the critical trials in order to see which interpretation they assigned to the ambiguous pronoun. Participants were consistent in the strategy adopted: 27 of the 30 participants rejected the target sentences 97% of the times overall, interpreting the ambiguous expression exophorically. The remaining three participants accepted the target sentences instead, consistently with the anaphoric interpretation.

We also compared participants' eye movements in the critical trials with their eye movements in the control trials that, as we said, could be differentiated on the basis of the type of exploration required to process them. To this purpose, we only considered the responses "false" to the critical trials in the statistical comparisons with the control trials. We analyzed the reaction time (RT) taken for the evaluation and the eye movement recordings. First, we checked the overall distribution of RTs for normality and excluded all RTs that exceeded 12,000 ms. Four responses were excluded on this basis, and 10 additional data points were excluded for technical reasons, so that the statistical comparisons included 1,717 data points overall.[4] We took the participants' ΔRT as the dependent variable, calculated by subtracting the real duration of the sentence from the overall RT taken to judge it, in consideration of the fact that the overall duration of the sentences varied a lot across trials, depending on the length of the words that were used and the way they combined to form each sentence. For example, a certain variation in the word length occurred in the names (e.g., "pentagon"/pen tagono/ vs. G/dʒi/), in the adjectives

(e.g., "piccolo"/ pikkolo/ vs. "blu"/blu/), in the predicates (e.g., "fa il poliziotto" (is a policeman) vs. "dorme" (sleeps)), or in the determiners ("il"/il/ vs. "nessun"/nes′sun/). A summary of sentence length and associated RTs by experimental condition is reported in Table 2.

Data were analyzed using mixed-effects analyses (Baayen, Davidson, & Bates, 2008) with random intercepts for subjects and items. A model was run in which the effects of two fixed variables were tested: the type of target used (shape, letter, digit), taking digits as the reference level; and the experimental condition (AMB critical trials, ONEq control trials, ALLq control trials), taking AMB as the reference level. No effect of the type of target was revealed, and the variable was thus removed from the model (it did not significantly contribute to the overall goodness of fit). Critical trials took significantly shortest ΔRTs than ALLq control trials, while no difference between critical trials and ONEq control trials was found. Fixed effects are reported in Table 3.

To better grasp what the participants were considering during the time course of the sentence, we plotted the proportions of fixations to the target quadrant over the other quadrants as a function of the sentence timeline (Fig. 3). In order for sentences of different length to be comparable: (a) timelines were obtained by dividing and stretching each sentence in 10 bins, and (b) sentences were centered (bin 0) on the beginning of the crucial interesting period, corresponding to the beginning of *the only one* in AMB critical sentences and the beginning of the verb phrase (VP) in the control trials. The "target quadrant" was the quadrant that could be identified soon after the end of the subject NP in all cases in which a noun referent was given in the NP, as it was the case for all critical trials and for ONEq control trials (e.g., the quadrant containing the squares for sentence [5] or the quadrant containing the fives for sentence [6]). In all the stimuli labeled ALLq, in which a target quadrant could not be identified after hearing the subject (as in [7]), we considered the proportion of looks to one of the four quadrants arbitrarily chosen.

What is evident from the graph is that different sentence types leads to different exploration patterns. In case of ALLq trials, participants' fixations were scattered equally across the four quadrants throughout the whole sentence (cf. the purple line in Fig. 3, flat at chance level). This exploration pattern is consistent with the fact that participants had to look around and consider all the quadrants to find examples that either verified or falsified the sentence. A different pattern seems to reveal in case of ONEq trials, in which participants looked more at the target quadrant from VP-onset (bin 0) and do not move out of it for the rest of the sentence (cf. the green line in Fig. 3). This was the quadrant that could be individuated after the end of the subject NP, like, for example, the quadrant

Table 2
Descriptive statistics for the different experimental conditions (Experiment 1): mean and standard deviation of sentence durations and associated response times

	Duration (ms)	RTs (ms)	ΔRTs (ms)
AMB	4,785 ± 306	5,210 ± 993	929 ± 987
ALLq	3,718 ± 425	5,263 ± 1,355	1,552 ± 1,279
ONEq	3,324 ± 398	4,303 ± 964	992 ± 870

Table 3
Fixed effects analysis on ΔRTs as a function of the experimental condition (Experiment 1)

	Estimate (ms)	*SE*	*t*-value	pMCMC
Intercept (AMB)	942	137	6.86	.0001
Condition: ALLq	634	100	6.32	.0001
Condition: ONEq	44	103	0.43	.6460

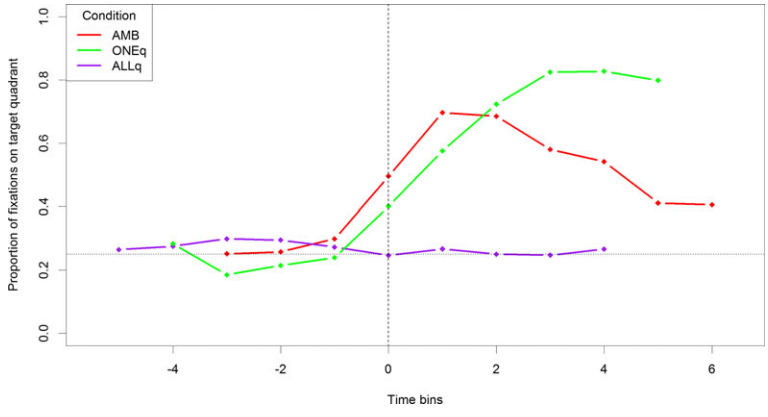

Fig. 3. Proportion of fixations to the target quadrant during the on-line processing of sentences in the different experimental conditions in Experiment 1. Each dot represents a timeline bin. The horizontal dotted line indicates the chance level (.25). The vertical dashed line represents the beginning of the critical period, corresponding to VP-onset in the control conditions (ONEq and ALLq) and to onset of the ambiguous segment *the only one* in the critical condition (AMB).

that contains the fives once the segment "All the fives" was heard in example (6). Crucially, the pattern of exploration that is revealed for the critical sentences (cf. the red line in Fig. 3) seems overall different from all the others. Before VP-onset, a preference for the target quadrant over the others is revealed: After the end of the NP, participants converged on the target quadrant (e.g., the quadrant that contains the squares once "The small square" is heard in example [5]). They kept fixating it during the ambiguous segment *the only one* (that approximately lasted until bin 3) but, differently from what happens for ONEq trials, the proportion of looks to this target starts decreasing in the final part of the sentence, signaling the fact that participants moved out of that quadrant and considered other quadrants as well before making a final decision on the sentence.

For the statistical analyses, we considered four interesting periods (IPs): IP1 corresponds to the duration of the whole phrasal subject; IP2 corresponds to the duration of the critical segment *the only one* in the critical trials and to a short pause between the subject NP and the VP in the control trials; IP3 corresponds to the duration of the whole VP; IP4 corresponds to the time to take a decision, starting from the end of the sentence until the response button is pressed. Each of the fixation events was assigned to the IP it landed in. We submitted our data to a series of logit mixed models (Jaeger, 2008) in which the likelihood of looking at the target quadrant was modeled as a function of the

experimental condition (AMB vs. ONEq vs. ALLq), for each of the four IPs. A summary of the statistical comparisons is provided in Table 4. AMB critical trials were taken as the reference level.

The proportion of looks to target in the critical trials differed from ONEq control trials in all the interesting periods considered, as shown in the last row in Table 4. Considering the period after VP-onset (i.e., IP3 and IP4 in the last columns of Table 4), the proportion of looks to the target in the critical trials was significantly different from both ONEq and ALLq control trials. In fact, during IP2, participants looked significantly more to the target quadrant, as opposed to the rest of the scenario, as indicated by the positive and significant intercept parameter; in contrast, during IP3, the rest of the scenario attracted more fixations than the target quadrant, as indicated by the negative and significant intercept parameter. Through inspection of the data (Fig. 3), we can see that the proportion of looks to the target quadrant in the critical trials starts decreasing before the end of *the only one*, which falls approximately at bin 3 in the graph, and then it falls middle ground between that of ONEq control trials (in which a target could be identified and, consequently, the proportion of looks to that target was very high) and of ALLq control trials (in which the proportion of looks was equally scattered across all quadrants). We will discuss these results in the next section.

2.3. Discussion

Several results are worth discussing at this point. Let us focus on the truth-value judgments first. As mentioned above, the critical sentences were true under the anaphoric (weak) interpretation and were false under the exophoric (strong) interpretation. This design was meant to contrast two hypotheses that made opposite predictions as far as truth-value judgments are concerned (cf. Table 1): the Principle of Charity predicted adults to select the weak interpretation, that is, the interpretation that made the ambiguous sentences true (in this case, this would be the anaphoric interpretation); the Principle of Maximizing Informativity predicted adults to select the strong interpretation, that is, the interpretation that is true in a narrower set of circumstances, thereby making the sentence false (in this case, this would be the exophoric interpretation). Results indicate that participants mostly judged the critical sentences false, thus interpreting the ambiguous expression *the only one* exophorically; also, the eye movement pattern that emerged in critical

Table 4
Summary of the results of the logistic regression models for each of the four interesting periods, taking AMB as the reference level (Experiment 1)

	IP1		IP2		IP3		IP4	
	Estimate	*p*	Estimate	*p*	Estimate	*p*	Estimate	*p*
Intercept (AMB)	−.67	.0004	.76	.0001	−.21	.0892	−.47	.0001
Condition: ALLq	−.46	.1249	−2.07	.0001	−.97	.0001	−.58	.0008
Condition: ONEq	−.81	.0098	−1.33	.001	1.39	.0001	1.28	.0001

trials was different from ONEq control trials, signaling the fact that the domain for the interpretation of the anaphora was not confined to a single quadrant once the target for the linguistic antecedent was identified. Overall, this result is compatible with the predictions of the Principle of Maximizing Informativity but contrary to the predictions of the Principle of Charity: Despite the fact that the anaphoric interpretation would have made the sentences true (and would have required the least effort in the exploration), adults seem to commit themselves to the strong reading and do not restrict the exploration of the scenario to the target quadrant that contains the linguistic antecedent, which would suffice to make the sentence true.

These findings do not replicate those obtained by Crain et al.'s in the Sentence Falsification task with adults; rather, they mirror the effect observed in the Truth-Value Judgment task with children in which the exophoric interpretation was favored. As argued by Foppolo et al. (2014), this difference might be task dependent: When adults are tested with the same task as children (in which a context is provided), they opt for the strong (exophoric) interpretation of the ambiguous expression, the one that conveys the most informative message.

By analyzing eye movements and RT measures, we found some other informative results. First, the analysis on RT measures revealed that participants were fast to make a decision in the critical trials and, in particular, they were significantly faster than ALLq control trials, in which no target quadrant could be identified and the scenario had to be explored thoroughly. This behavior alone might be surprising, given the fact that an ambiguity was present in the critical sentences and, allegedly, this should slow the processing down. Second, the fixation pattern plotted in Fig. 4 reveals interesting analogies and differences across the experimental conditions. Recall that the Principle of Charity predicted a similarity in the pattern of exploration between AMB critical trials and ONEq control trials, while the Principle of Maximizing Informativity predicted a difference between them (cf. Table 1). In ALLq control trials, participants never focused on one specific quadrant but their fixations were scattered throughout the whole scenario, presumably in the attempt of verifying or falsifying the sentence. In ONEq control trials, participants looked around until they found the target quadrant and kept focusing on that quadrant until the end of the sentence to see if the predicate was verified or falsified within that quadrant. Crucially, they almost never moved out of that quadrant once it was identified. In the critical AMB trials, a different pattern of exploration emerged: At first, participants looked around until they found the target quadrant and focused on that quadrant for a while but then, differently from what they did in ONEq control trials, they started to move out of the relevant quadrant soon after they heard *the only one*. In fact, RT and eye movement data seem to exclude that participants entertained the anaphoric interpretation of the ambiguous expression in a preliminary stage of processing: If they did, then the fixation pattern in the critical trials immediately after the ambiguous pronoun should have resembled that of ONEq control trials, but the analyses do not support this conclusion.

Eye movement data also show that participants started to move out of the relevant quadrant before the end of the predicate. This behavior is particularly revealing about the

on-line sentence comprehension process, independently of the presence of the anaphora. In fact, participants initiated their visual search outside the target quadrant, presumably in search of referents that could probe or disconfirm the validity of the sentence, before actually "knowing" for sure what to look for, provided that the predicate under consideration has not been fully revealed at the beginning of IP3 yet. To make sense of participants' behavior, we offer the following explanation. In IP2, the critical sentences contained the ambiguous expression *the only one* that, besides being underspecified and thus in need of a referent, introduces a presupposition-like implication. Several theories of the meaning of *only* have been suggested in the linguistic literature. Independently of the semantic account adopted, all agree that a sentence containing *only* like (8) conveys the two propositions in (9):

(8) *Only* John smokes

(9) (a) John smokes (*prejacent implication*)
(b) No one else, except for John, smokes (*exclusive implication*)

While the exclusive implication (9b) is pretty clearly an entailment of sentence (8), divergence exists about the status of the prejacent implication (9a), that is analyzed as an entailment, a presupposition, or a conversational implicature in different accounts (cf. Ippolito, 2008 and references therein). Independently of the semantic analysis that we adopt, a question that goes beyond the purpose of this paper, we believe that it is exactly the "implication" in (9a) that the subjects in our experiment exploit immediately as a cue to speed up the processing of the ambiguous sentences. In particular, once the segment with *only* was heard, subjects could anticipate the predicate under discussion by exploiting the cue of the "prejacent implication." For example, when hearing the segment "The small square is the only one that" in a scenario like Fig. 1, subjects could first converge on the target (the quadrant with the squares, and the small square in particular) and then started to anticipate what came next in the sentence on the basis of what the small square was in fact doing (i.e., being a policeman). On the basis of this information, they could immediately start moving around to check for the validity of the statement; that is, they could start looking around for other policemen. This would explain why subjects started to move out of the target quadrant before the end of the predicate and why they were so fast in giving a judgment. Note that this anticipation was only possible in case of the critical trials; in all the other cases, to know what the sentence was about, one needed to wait until the whole predicate unfolded.

This interpretation of our data fits well with interactive accounts of sentence processing, according to which the parser is not modular but makes use of multiple cues in analyzing a complex linguistic input incrementally. Our results corroborate and extend previous findings in which anticipatory effects on the exploration pattern of a visual scenario have been widely documented for purely linguistic cues (like case-marking, cf. Kamide et al., 2003) or contextual cues in the visual input (a.o. Knoeferle, Crocker,

Scheepers, & Pickering, 2005; Knoeferle & Crocker, 2006; Sedivy et al., 1999; Tanenhaus et al., 1995).

As for our first experimental question, the results of Experiment 1 speak against the Principle of Charity, while they seem compatible with the Principle of Maximizing Informativity, as formulated above (see section 2.1.3): Adults, in fact, did not seem to select the weak alternative in solving privative ambiguities but opted for the strong reading (the exophoric, in this case). Also, RT and eye-tracking measures showed that the participants were very fast in making a decision and made use of all the linguistic and non-linguistic cues to speed up their search and resolve the ambiguity.

Indeed, this result alone is not particularly convincing from a theoretical point of view: What seems to emerge from Experiment 1 is that adults rely on considerations about informativity in solving semantic ambiguity and, in particular, follow the Principle of Maximizing Informativity. However, further experimental evidence is needed to corroborate this hypothesis. This is the purpose of the second experiment, to which we now turn.

3. Experiment 2: Do adults maximally commit themselves to the strong alternative?

In Experiment 1, we found that participants interpreted *the only one* exophorically and overwhelmingly judged the ambiguous sentences false in a scenario in which they were that made them true under the anaphoric interpretation. As we said, this finding seems compatible with the hypothesis that adults rely on considerations about informativeness and select the stronger among alternatives, favoring the Principle of Maximizing Informativity over the Principle of Charity. As we anticipated, this is not enough evidence to fully establish that adults solve privative ambiguity relying on considerations about informativeness. To provide further support for the Informativeness Principle, one needs to show that participants' judgments revert when the polarity of the context that regulates informativeness and entailment relations reverts, too. To this purpose, in Experiment 2 we reversed the pattern of entailment between the alternative interpretations of the critical ambiguous sentences. This was obtained by simply inserting negation in the original sentences used in Experiment 1, so that the ambiguous expression became *not the only one*. Under negation, the exophoric interpretation, the one that was the strong interpretation in Experiment 1, becomes the weak interpretation, the one that requires fewer commitments for the truth of the sentence. Back to our initial example, by uttering "The tall girl is *not the only one* that is wearing a hat" in a context of five boys and five girls in which one boy and one girl are wearing hats, one would say something true under the exophoric interpretation (the tall girl is not the only *individual* in the domain wearing a hat, provided that a boy is wearing a hat too), but it would say something false under the anaphoric interpretation (the tall girl is indeed the only *girl* in the domain of discourse that is wearing a hat, provided that no other girl, except for the tall one, is wearing a hat). This demonstrates that, under negation, the more informative interpretation is the anaphoric one. If the Informativeness Principle is

at stake and adults maximally commit themselves to the strong alternative, as advocated by the Principle of Maximizing Informativity (compatible with the findings of Experiment 1), participants should interpret the ambiguity anaphorically under negation. This means that they would focus on the single quadrant that contains the individuals of the same individual-level category of the linguistic antecedent of the anaphora, paralleling ONEq control trials this time. If, conversely, adults do not rely on considerations about informativeness, they might extend their domain of evaluation to include the whole scenario and solve the ambiguity exophorically, as they did in Experiment 1, albeit judge the critical sentences true this time. The expectation in this case is that their eye movements would mirror what has been already found in Experiment 1; that is, participants should explore the whole scenario and, in particular, they should behave differently from ONEq control trials in this case, too.

3.1. Methods

3.1.1. Participants

Twenty-one students of the University of Milano-Bicocca participated in the experiment and received credits for their participation. All subjects were native speakers of Italian, had normal or corrected-to-normal vision, and were naive with respect to the goals of the experiment.

3.1.2. Material

The same materials of Experiment 1 were used in this second experiment. The only manipulation was that we turned the 30 critical ambiguous sentences into negative sentences by simply inserting negation in the ambiguous segment: *è l'unico/a* turned into *non è l'unico/a* (*is not the only-one*$_{[+MASC/+FEM]}$). Everything else was identical to Experiment 1. In particular, no changes were made in the visual scenarios and in the control trials. To exemplify, the target sentence (5) in Experiment 1, here repeated as (10) for convenience, was turned into (11) in Experiment 2:

(10) Il quadrato piccolo è l'unico che fa il poliziotto
the square small is the only-one that makes the policeman
The small square is the only one that is a policeman

(11) Il quadrato piccolo non è l'unico che fa il poliziotto
the square small not is the only-one that makes the policeman
The small square is not the only one that is a policeman

Crucially, (11) had to be evaluated with respect to the same scenario used in Experiment 1 for sentence (10), that is, Fig. 1. As in the previous experiment, the critical sentences were ambiguous between an exophoric and an anaphoric interpretation. In this case, though, the presence of negation reversed the entailing pattern, with the result that

the anaphoric interpretation entails the exophoric interpretation this time, as detailed in (12):

(12) (a) The small square is not the only *square* that is a policeman *anaphoric* (false)
(b) The small square is not the only *thing/shape* that is a policeman *exophoric* (true)
(c) entailing pattern: (a) entails (b)

For this reason, under the hypothesis that the entailment pattern of alternatives is taken into account in solving privative ambiguities, the final truth-value judgment should revert, too: Sentence (11) should be judged false if *the only one* was interpreted anaphorically, as in (12a), given that, *among the squares*, the small one is in fact the only one that is a policeman; conversely, it should be judged true if *the only one* is interpreted exophorically, as in (12b), given that, *among all the things*, the small square is not the only policeman; in fact, a triangle and a circle are policemen, too (cf. Fig. 1).

3.1.3. Predictions

If participants rely on the Principle of Maximizing Informativity, as the results of Experiment 1 seem to show, there is only one result that would be compatible with this hypothesis: The answer FALSE is still expected, but the anaphoric interpretation adopted in this case. As a consequence, the following pattern in participants' eye movements is predicted: The critical ambiguous sentences (AMB) should pattern like ONEq control trials. For the sake of clarity, we summarize these predictions in Table 5.

3.1.4. Technical details

The same settings and apparatus used in Experiment 1 were employed.

3.1.5. Procedure

The same procedure used in Experiment 1 was adopted.

3.2. Results

As in Experiment 1, we first checked subjects' answers on control trials for accuracy. Participants scored 96% correct (mean accuracy), with a minimum score of 92% correct (*SD* 3%). One subject was excluded on the basis of bad performance (less than 2 *SD* below mean accuracy). We further excluded 29 incorrect answers over the 691 provided on control trials (about 4%). We also checked the overall distribution of RTs for normality and excluded all RTs that exceeded 8,000 ms. Five additional responses were excluded on this basis.[5]

Then, we analyzed the truth-value judgments provided by our participants in case of the critical trials, to see what was the interpretation assigned to the ambiguous sentences. As in Experiment 1, participants were consistent in the strategy adopted and an overall preference for one type of judgment emerged; in this case, 18 of the 20 participants accepted the target sentences more than 99% of the times overall, while the remaining two participants rejected the target sentences.

Table 5
Predictions of the Principle of Maximizing Informativity in the two experiments for truth-value judgments, interpretation, and exploration pattern

	Experiment 1 (*the only one*)	Experiment 2 (*not the only one*)
Truth-value judgment	FALSE	FALSE
Interpretation of *one*	exophoric	anaphoric
Exploration pattern	AMB $\neq$ ONEq	AMB = ONEq

Finally, we analyzed the RT taken for the evaluation and the eye movement recordings. Considering the fact that the judgment provided by the participants reflected the interpretation associated with the ambiguous expression (anaphoric vs. exophoric) and given that this, in turn, might be associated with a different pattern of exploration (paralleling ONEq and ALLq control trials, respectively), we only considered the responses "true" to the critical trials in the statistical comparisons with the control trials. Analyses included 1154 data points overall. A summary of sentence length and associated RTs by experimental condition is reported in Table 6.

As done in the previous experiment, data were analyzed using mixed-effects analyses (Baayen et al., 2008) with random intercepts for subjects and items, taking participants' ΔRT (as defined above) as the dependent variable. Fixed effects are summarized in Table 7.

As in Experiment 1, critical trials took significantly the shortest ΔRTs than ALLq control trials, which required the exploration of the whole scenario to be evaluated. Interestingly, though, in Experiment 2 the time taken to make a decision on critical trials was also significantly shorter than ONEq control trials, which required the exploration of one single quadrant. As before, no effect of the type of target was found. Analogously to Fig. 3 in the previous experiment, Fig. 4 shows the proportions of fixations to the target quadrant over the other quadrants as a function of the sentence timeline.

Table 6
Descriptive statistics for the different experimental conditions in Experiment 2: Mean and standard deviation of sentence durations and associated response times (ΔRTs)

	Duration (ms)	RTs (ms)	ΔRTs (ms)
AMB	4,785 ± 306	5,512 ± 467	724 ± 438
ALLq	3,718 ± 425	4,874 ± 764	1,156 ± 640
ONEq	3,324 ± 398	4,208 ± 600	893 ± 482

Table 7
Fixed effects analysis on ΔRTs as a function of the experimental condition (Experiment 2)

	Estimate (ms)	*SE*	*t*-value	pMCMC
Intercept (AMB)	726	75	9.69	.0001
Condition: ALLq	446	73	6.11	.0001
Condition: ONEq	172	76	2.28	.0074

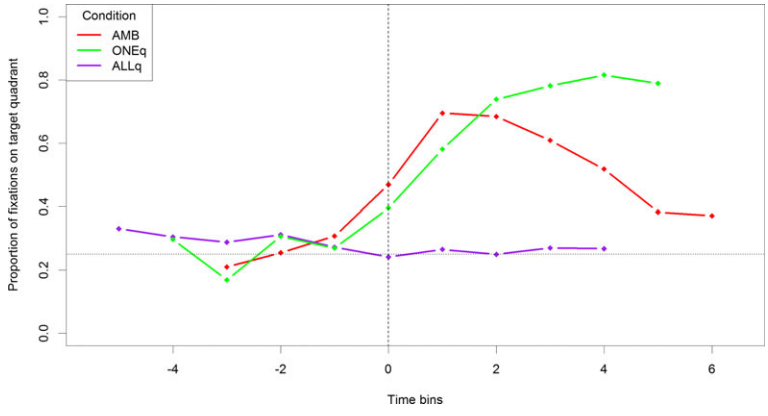

Fig. 4. Proportion of fixations to the target quadrant during the on-line processing of sentences in the different experimental conditions in Experiment 2. Each dot represents a timeline bin. The horizontal dotted line indicates the chance level (.25). The vertical dashed line represents the beginning of the critical period, corresponding to VP-onset in the control conditions (ONEq and ALLq) and to onset of the ambiguous segment *the only one* in the critical condition (AMB).

The eye movement pattern that is found in Experiment 2 is very similar to the one observed in Experiment 1. As before, in case of ALLq control trials, participants' fixations were scattered equally across the four quadrants throughout the whole sentence, as it was expected provided that no target quadrant could be identified and the whole scenario should be checked thoroughly before providing a truth-value judgment. In case of ONEq control trials, instead, participants' fixations converged to the target quadrant after hearing the phrasal subject and then clustered on that single quadrant until the end of the sentence, when a judgment was provided. In case of the AMB critical trials, participants' fixation pattern is middle ground between the two control trials: participants first focused on the target quadrant (e.g., the quadrant with the squares when processing the subject of sentence [11]) but, differently from ONEq control trials, they moved out of that quadrant after hearing *not the only one* and considered other quadrants as well, as found in the previous experiment. Turning to the statistical analysis of fixations, we segmented the time course of the recording into four different interesting periods, as we did in the previous experiment. We then submitted our data to a series of logit mixed models in which the likelihood of looking at the target

Table 8
Summary of the results of the logistic regression models for each of the four interesting periods, taking AMB as the reference level (Experiment 2)

	IP1		IP2		IP3		IP4	
	Estimate	*p*	Estimate	*p*	Estimate	*p*	Estimate	*p*
Intercept (AMB)	−.76	.0001	.66	.0001	−.42	.0003	−.35	.0169
Condition: ALLq	−.47	.1223	−1.91	.0001	−.66	.0002	−.68	.0014
Condition: ONEq	−.65	.0397	−1.26	.0001	1.55	.0001	1.15	.0001

quadrant was modeled as a function of the experimental condition for each of the four interesting periods considered. A summary of the statistical comparisons is reported in Table 8. As before, critical trials (AMB) were taken as the reference level.

Also in this case, a difference between critical AMB trials and ONEq control trials emerged in all the IPs considered, as it is evident in the last line of Table 8. As before, during IP2 participants looked significantly more to the target quadrant, as opposed to the rest of the scenario (as indicated by the positive and significant intercept parameter), while they looked significantly more to the rest of the scenario during IP3, as indicated by the negative and significant intercept parameter.

3.3. Discussion

In Experiment 2, we found that participants interpreted *the only one* exophorically, as they did in Experiment 1, and, as before, they extended their domain of evaluation to the whole scenario also once a target quadrant was identified. Crucially, though, the truth-value judgment provided in Experiment 2 was TRUE, while it was FALSE in Experiment 1. This is consistent with the fact that the entailment pattern between the alternative readings of the ambiguous pronoun was reversed in this case, due to the presence of negation. As a consequence, the exophoric interpretation was the weak reading this time, the one that required less commitments for its truth, while it was the strong reading in Experiment 1.

The eye movement data were revealing, too. First of all, participants' eye movements in Experiment 2 were very similar to the one observed in Experiment 1. This was expected in case of control trials, considering the fact that these were exactly the same in the two experiments. As before, participants' fixations were scattered equally across the four quadrants throughout the whole sentence in case of ALLq control trials, while they clustered on the single-target quadrant in case of ONEq control trials, once the target was identified. The crucial finding, though, is that even the pattern of exploration in case of AMB critical trials was very similar in the two experiments; in both studies, participants end up considering the whole scenario and solved the ambiguity exophorically, independently of truth-conditional considerations. As in the previous study, participants first focused on the target quadrant (containing the individuals of the same individual-level category of the linguistic antecedent) and then started to move out of that quadrant before the end of the predicate, after hearing the segment containing *not the only one*. As discussed before, they did so presumably in search of other individuals performing the same action as the target referent. As for Experiment 1, this effect might be explained by the "prejacent implication" associated with *only*, that resists negation (Geurts & van der Sandt, 2004).

Turning to RTs, it is interesting to note that the critical trials in this second experiment were significantly faster than both control conditions, even those that required the exploration of a single quadrant (ONEq). We think that this result might depend on the fact that, as before, we calculated the ΔRT by subtracting the duration of the whole sentence from the time taken to provide a response. As a result, the duration of the control

sentences was identical in the two studies, but the duration of the critical sentences was slightly longer in Experiment 2 because of the *non* segment. This might have allowed for extra time to exploit the cue associated with *only*, provided that the ΔRT for the critical sentences, on which the analyses were conducted, was at net of this extra "reasoning" time.

4. General discussion

In this paper, we investigated the interpretation of privative ambiguities in two eye-tracking experiments with adults. By means of the Visual World Paradigm, we asked participants to judge sentences like "The small square is the only one that is a policeman" that are ambiguous between the *anaphoric* (the small square is the only *square* that ...) and the *exophoric* interpretation (the small square is the only *individual* that ...). Crucially, these alternative interpretations are not truth-conditionally independent from one another, one interpretation being stronger (or more informative) than the other. Two widespread assumptions have been considered theoretically to account for cases like this, in which an entailment relation occurs between alternatives: (a) Adults prefer stronger readings over weaker ones, following what we dubbed a Principle of Maximizing Informativity (Principle of Pragmatic Strengthening; Maximal Commitment strategy); (b) adults choose the interpretation that makes a sentence true whenever possible, being charitable or making the least commitments possible (Principle of Charity; Minimal Commitment strategy). Both principles are based on a more general principle according to which adults rely on considerations about informativity in resolving ambiguity, following the Informativeness Principle. As we said, though, a direct investigation of the role of informativeness in processing privative ambiguities is missing in the traditional psycholinguistic literature on anaphora resolution in adults (cf. Altmann, 2001 for an overview and references therein), with the exception of the study by Crain et al. (1994) discussed above.

To fill this gap, we examined participants' eye movements during the on-line processing of privative ambiguities. As we said, two alternative interpretations might be assigned to *the only one*: the anaphoric and the exophoric. Crucially, these two interpretations map onto different exploration patterns to restrict the domain of quantification: Either the whole scenario is explored, or the exploration is restricted to the set of individuals of the same individual-level category of the linguistic antecedent once a target quadrant is identified. These options were matched to two control conditions (ONEq vs. ALLq) and refer to the exophoric and the anaphoric interpretation of the pronoun, respectively.

Starting from the study by Crain et al., Experiment 1 tested which interpretation was favored by adults and which kind of exploration they adopted in order to judge the ambiguity. We found that subjects overwhelmingly solved the ambiguity exophorically by enlarging the domain of evaluation so as to include other sets of entities. In terms of the Informativeness Principle, the results of Experiment 1 were overall inconsistent with the Principle of Charity while supporting the Principle of Maximizing Informativity, provided that the exophoric interpretation was the strong one in that case. This, however, did not

constitute enough evidence in favor of the claim that adults rely on truth-conditional considerations and prefer strong readings by default. In fact, participants might have interpreted *the only one* exophorically in Experiment 1 not because they were maximally committed to the most informative interpretation, but for other reasons as well. To test whether considerations about informativity were the key to participants' behavior in the first experiment (as maintained by the Informativeness Principle), we manipulated informativity by inserting negation in Experiment 2 so as to revert the polarity of the sentence. In particular, negation reverted the ordering of the alternatives on the scale of informativity so that the exophoric interpretation became the weak interpretation, the one that rendered the sentence true. If participants were guided by considerations about informativeness and followed the Principle of Maximizing Informativity, they would shift to the anaphoric interpretation this time, this being the strong one. What we found, instead, was that participants enlarged their domain of evaluation so as to include the whole scenario, exactly as they did in Experiment 1: In so doing, they kept interpreting *the only one* exophorically, independently of the fact that this interpretation was the weak one this time, the one that conveyed the least informative message.

On the basis of the conjoined results of the two experiments presented here, we cast doubts on the idea that truth-conditional properties of alternatives play a role in processing ambiguity, contra the Informativeness Principle, and supporting Geurts (2000)'s theoretical aversion to preferences based on logical strength. In fact, the participants in our two studies did not commit themselves to the strong interpretation to obey a general Principle of Maximizing Information. What they did, instead, was to extend the exploration outside the target quadrant once it was identified. We might call this a general *Principle of Maximal Exploitation*: When a context is provided, adults maximally exploit it (and do so as soon as they can). This principle seems to operate independently from truth-conditional considerations and readily accounts for the fact that participants interpreted *the only one* exophorically in our studies, independently of the fact that this interpretation was the weak one under negation.

In accordance with interactive accounts of sentence processing, our results are compatible with many well-attested results presented in the literature on sentence comprehension in which different contextual sources were shown to influence processing. Just to mention a few, visual context was proved to influence the processing of syntactically ambiguous structures incrementally in the pioneering study by Tanenhaus et al. (1995) in which participants were asked to move real objects in a display following instructions like "Put the apple on the towel in the box" when one or two apples were displayed. In that study, the presence of two objects induced subjects to interpret the prepositional phrase *on the towel* as a modifier of the noun phrase *the apple*. Analogously, preceding linguistic context was shown to influence the processing of syntactically ambiguous sentences like "The burglars blew open the safe with the rusty lock," in which a preference for low attachment of the prepositional-phrase modifier (the safe with the rusty lock) was recorded in a context in which two safes were mentioned (Altmann & Steedman, 1988).

However, maximally exploiting the context does not mean that the listeners' attention is not selective or irrespective of relevant features or cues provided within the context.

Quite the contrary, in fact: Maximally exploiting the context means exactly to maximize all the information that the context provides, this information being explicit (as it is the case with visual competitors) or implicit (as it is the case with many hidden cues linked to the semantics of the words used in the sentence). In our study, for example, we found that, in both experiments, participants started moving out of the relevant quadrant, presumably in search of other individuals performing the same action as the main subject, before the end of the predicate. We attempted to account for this fact in terms of the "prejacent implication" related to *only*, according to which participants could anticipate the predicate under discussion and used that cue immediately to speed up their visual search. Again, analogous effects are documented in a wide range of studies reported in the literature in which selective attention to a partition of the visual domain is reported. For example, Altmann and Kamide (1999) showed how lexical information associated with the verb could be used to restrict the visual domain to objects that might function as possible arguments of the verb. Analogously, Chambers, Tanenhaus, Eberhard, Filip, and Carlson (2002) presented evidence that linguistic and non-linguistic information sources are combined to constrain referential interpretation (see also Sedivy et al. [1999] on prenominal adjectives) and Knoeferle et al. (2005) argued for a theory of on-line sentence comprehension that exploits a rich inventory of semantic categories, showing eye-tracking evidence that visual representation of events can affect thematic-role assignment incrementally.

Also, the Principle of Maximal Exploitation might be seen as the downside of the well-known Principle of Parsimony according to which adults should prefer the reading that carries fewer unsupported presuppositions (Altmann & Steedman, 1988; Crain & Steedman, 1985): (a) When no context is provided, as in the original Crain et al.'s study with adults, participants avoid postulating additional information, as the Principle of Parsimony predicts, and (b) when a context is provided instead, adults maximally exploit what they get. This is revealed by the participants' eye movements in the two studies reported here or in our questionnaire study cited above (Foppolo et al., 2014) in which the richness of the context affected how participants extended the domain of quantification to solve privative ambiguity. This conjoined result means that adults pinpoint all the relevant cues available in the visual and linguistic context and exploit them as early as they can, as the Principle of Maximal Exploitation predicts, in accordance with interactive and incremental accounts of sentence processing. Contrary to the Informativeness Principle, traditionally ensued from semantic/pragmatic theorizing, our findings suggest that considerations about strength and informativeness do not take part in the resolution of privative ambiguities in context.

Acknowledgments

This work was (partially) supported by funds from the European Science Foundation (Collaborative Research in Experimental Pragmatic, EURO-XPRAG) to the first author and by a VIDI grant from the Netherlands Organization for Scientific Research (NWO) to

Andrea Gualmini. We want to thank the three anonymous reviewers for their useful suggestions that contributed to shape the final version of this manuscript. We also thank Patrice Rusconi for insightful discussion and Ailis Cournane for her help in the preparation of the stimuli.

Authors' contributions are as follows: Francesca, Luisa, and Andrea originally conceived the experiment, with subsequent Marco's contribution; Francesca, Luisa, and Andrea contributed in creating the stimuli; Francesca implemented the experiment script and collected the data; Marco analyzed the data with Francesca's contribution. All authors drafted the paper, with critical revisions from Francesca and Marco. Francesca supervised the project.

Notes

1. Deriving the implicatures *some but not all* in the antecedent of a conditional would make a weaker claim, making the conditional true if the girl ate all of the cookies but did not gain weight (Chierchia, 2006; Kadmon & Landman, 1993).
2. In the present discussion, we leave aside the phenomenon of Scalar Implicatures, which have been extensively investigated experimentally. However, we believe that the mechanisms involved in the interpretation of scalar terms might be somehow different from those involved in anaphora resolution of the type investigated here; for example, the derivation of Scalar Implicatures is tightly linked to pragmatic principles like the Gricean maxim of Quantity, which are not necessarily involved in the process studied in this paper.
3. One of the reviewers raised a potential problem for our stimuli for languages that, like English, contain a reference to a "human" or "human-like" entity within the expression *the only one*. As the reviewer pointed out, the English expression *the only one* shares the segment *one* with words like some*one*, every*one*, and no *one* and this might force a reading in which the human-like emoticon is separated from the abstract object to which it is attached and considered separately in the evaluation. Differently from English, the Italian null form that was used in our study (i.e., *l'unico*/*l'unica*) does not contain any reference to a human *one*, and it is not contained in words like *tutti* (all, everyone), *nessuno* (no one), *qualche*, or *qualcuno* (some, someone). Moreover, participants were specifically trained to associate the emoticons to the abstract objects and to interpret them as the representation of the actions being performed.
4. The same data set was obtained when checking normality on ΔRTs.
5. The same data set was obtained when checking normality on ΔRTs.

References

Abusch, D. (1993). The scope of indefinites. *Natural Language Semantics*, *2*, 83–135.

Altmann, G. T. M. (2001). The language machine: Psycholinguistics in review. *British Journal of Psychology*, *92*, 129–170.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264.

Altmann, G. T. M., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*, 191–238.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*, 390–412.

Blackburn, S. (1994). *The oxford dictionary of philosophy* (p. 62). Oxford, UK: Oxford University Press.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language*, *47*(1), 30–49.

Chierchia, G. (2006). Broaden your views: Implicatures of domain widening and the "logicality" of language. *Linguistic Inquiry*, *37*(4), 535–590.

Chierchia, G., Fox, D., & Spector, B. (2013). The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In P. Portner, C. Maienborn, & K. von Heusinger (Eds.), *Handbook of semantics*. Berlin: Mouton de Gruyter.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real–time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84–107.

Crain, S., & McKee, C. (1985). The acquisition of structural restrictions on anaphora. In S. Berman, J-W. Choe, & J. McDonough (Eds.) *Proceedings of NELS 15* (pp. 94–110). Amherst, MA: GLSA.

Crain, S., Ni, W., & Conway, L. (1994). Learning, parsing and modularity. In C. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing*. Hillsdale, NJ: LEA.

Crain, S., & Steedman, M. (1985). On not being led up the garden path: The use of context by the psychological parser. In D. Dowty, L. Karttunnen, & A. Zwicky (Eds.), *Natural language parsing*. Cambridge, UK: Cambridge University Press.

Dalrymple, M., Kanazawa, M., Kim, Y., Mchombo, S., & Peters, S. (1998). Reciprocal expressions and the concept of reciprocity. *Linguistics and Philosophy*, *21*, 159–210.

Davidson, D. (1984). *Inquiries into truth and interpretation*. Oxford, UK: Clarendon Press.

Foppolo, F., Meroni, L., Marelli, M., & Gualmini, A. (2014). The only one strategy in context. *Lingua*, *145*, 266–275.

Geurts, B. (2000). Buoyancy and strength. *Journal of Semantics*, *17*, 315–333.

Geurts, B. (2002). Donkey business. *Linguistics and Philosophy*, *25*, 129–156.

Geurts, B. (2009). Scalar implicature and local pragmatics. *Mind and Language*, *24*, 51–79.

Geurts, B., & Pouscoulous, N. (2009). Embedded implicatures. *Semantics and Pragmatics*, *2*(4), 1–34.

Geurts, B., & van der Sandt, R. (2004). Interpreting focus. *Theoretical*. *Linguistics*, *30*, 2–44.

Horn, L. (1996). Exclusive company: "Only" and the dynamics of vertical inference. *Journal of Semantics*, *13*, 1–40.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171.

Ippolito, M. (2008). On the meaning of only. *Journal of Semantics*, *25*(1), 45–91.

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.

Kadmon, N., & Landman, F. (1993). Any. *Linguistics and Philosophy*, *16*(4), 353–422.

Kamide, Y., Scheepers, C., & Altmann, G. T. (2003). Integration of syntactic and semantic information in predictive processing: Cross-linguistic evidence from German and English. *Journal of Psycholinguistic Research*, *32*(1), 37–55.

Knoeferle, P., & Crocker, M. W. (2006). The coordinated interplay of scene, utterance, and world knowledge: Evidence from eye tracking. *Cognitive Science*, *30*(3), 481–529.

Knoeferle, P., Crocker, M. W., Scheepers, C., & Pickering, M. J. (2005). The influence of the immediate visual context on incremental thematic role-assignment: Evidence from eye-movements in depicted events. *Cognition*, *95*(1), 95–127.

Krifka, M. (1996). Pragmatic strengthening in plural predications and donkey sentences. In T. Galloway & J. Spence (Eds.), *Proceedings of the Sixth Conference on Semantics and Linguistic Theory* (pp. 136–153). Ithaca, NY: Cornell University.

Longhurst, E. (2006). *ExpBuilder*. Computer program.

Meyer, M. C., & Sauerland, U. (2009). A pragmatic constraint on ambiguity detection. *Natural Language & Linguistic Theory*, *27*, 139–150.

Quine, W. V. O. (1964). *Word and object*. (Vol. 4), Cambridge, MA: MIT Press.

Reinhart, T. (1997). Quantifier scope: How labor is divided between QR and choice functions. *Linguistics and Philosophy*, *20*, 335–397.

Sedivy, J., Tanenhaus, M., Chambers, C., & Carlson, G. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition*, *71*, 109–147.

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Winter, Y. (2001). Plural predication and the Strongest Meaning Hypothesis. *Journal of Semantics*, *18*, 333–365.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. List of experimental stimuli used in Experiment 1 and 2.