

# 1 Experiment 1

We used a sentence-picture matching task where participants saw a sentence and had to match the sentence with one of two pictures (modelled after Raffray & Pickering 2010 and Maldonado et al 2017, among others).

In experimental trials, the sentence always involved mixed adjectival predicates (either positive or negative) and was ambiguous between a collective and a distributive reading. Each sentence was presented with two out of three possible pictures: (a) a **foil** picture, that made both readings of the sentence false, (b) a **weak** picture, that made only the *weak* reading of the sentence true (whether the weak reading is the collective or the distributive one depends on the polarity of the adjective, see ??), and (c) a **‘blur’** picture, in which the relevant information was blurred so participants could not see it. Specific arrangements between pictures and sentences gave rise to two experimental items: primes and targets (see illustration in Figure 1).

There were two types of **primes trials**. *Collective primes* displayed a foil picture and a weak *collective* picture; *distributive primes* displayed a foil picture and a weak *distributive* picture. The nature of the weak reading varies depending on the polarity of the adjective. Thus, collective primes always involve positive adjectives (e.g. heavy), whereas distributive primes involve negative adjectives (e.g. light), see Table 1. Primes were designed to force one specific sentence interpretation: Participants should click in whatever image made the sentence true, and they would then be constrained to access the reading instantiated by the weak picture. Participants would access the collective reading of the sentence in collective primes, and the distributive reading in distributive primes.

PRIME/TARGET CONDITION	
Collective	Distributive
Positive adjective (heavy, noisy, expensive)	Negative adjective (light, quiet, cheap)

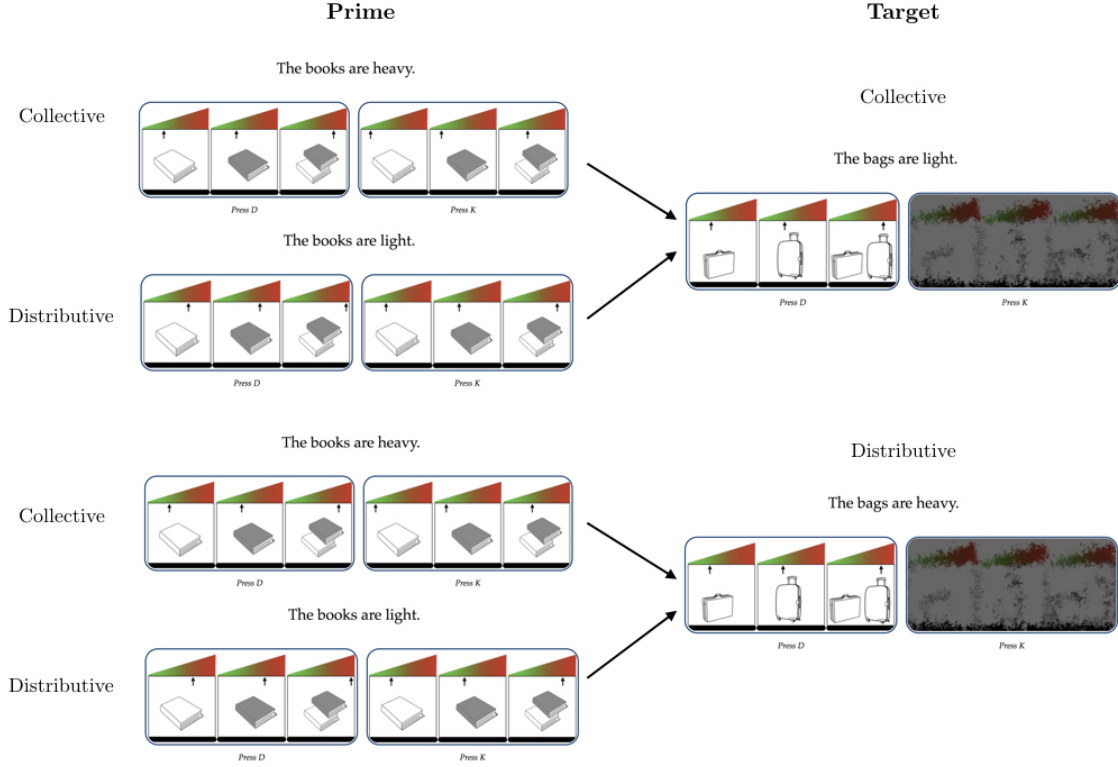
**Table 1:** Adjective polarity in each prime/target condition.

**Target trials** could also be either *collective* or *distributive*, depending on whether the sentence instantiate a positive or a negative adjective. Unlike primes, targets displayed a weak picture together with a ‘blur’ picture. Participants were instructed to select the ‘blur’ option if they did not feel that the overt picture sufficiently captured the sentence meaning (we modelled the “covered picture” method on Huang et al. 2013).

	COLLECTIVE TARGET (Positive adjective)	DISTRIBUTIVE TARGET (Negative adjective)
OVERT <i>weak</i> PICTURE	Collective reading	Distributive reading
BLUR PICTURE	Distributive reading (Collective reading)	Collective reading (Distributive reading)

**Table 2:** Readings compatible with each picture in target conditions.

Thus, we expected participants to click on the weak picture if they had accessed the weak reading (collective or distributive depending on the adjective polarity), and opt for the ‘blur’ option if they had accessed the strong reading (i.e. the reading that entails the other readings).



**Figure 1:** Illustration of experimental trials.

Table 2 summarises the readings compatible with each image for each target condition. Target trials immediately followed prime trials, and the four possible prime-target combinations were instantiated in the experiment. After being biased towards the collective reading in a *collective prime*, participants were expected to select more often a picture compatible with this collective reading in targets. In collective targets, this picture was the overt picture, whereas in distributive targets it was the ‘blur’ option. Conversely, distributive priming would be observed whenever, after a distributive prime, participants select more often the ‘blur’ picture in collective targets, and the overt weak picture in distributive targets. Thus, a semantic priming effect would be shown by an interaction between Prime and Target conditions in the image selection in targets.

## 1.1 Methods

### 1.1.1 Participants

Fifty four participants were recruited using Amazon Turk (Female=32). All of them reported English as their native language.

### 1.1.2 Materials

Each trial involved a sentence presented with two pictures. Participants had to match the sentence to one of the pictures. Sentences in experimental trials were constructed using the following frame: *The [plural exemplar] are [predicate]*. Predicates were relative adjectives in their positive or negative form, depending on the target/prime condition. They were selected among three

possible predicate pairs, depending on the scale dimension: light/heavy, cheap/expensive and noisy/quiet. Exemplars varied depending on the predicate pair. Predicates from the same dimension (i.e. each pair) had associated two different exemplars: ‘bags’ and ‘books’ for light/heavy; ‘cars’ and ‘watches’ for cheap/expensive and ‘birds’ and ‘instruments’ for noisy/quiet.

Pictures in experimental trials could belong to one of three categories: foil, weak and ‘blur’. Foil and weak pictures displayed three scales, which could measure degrees of various dimensions (weight, price or sound intensity). *How much detail should I provide here?* Values at the green portion of the scale represented low degrees; values at the red portion represented high degrees. Colour codes were provided to avoid confusion about the cut-off degree in the scale. The two left-most scales contained two different tokens of the exemplar named in the sentence (e.g. two different bags); the right-most scale had the two tokens together. ‘Blur’ pictures were constructed in such a way that participants could see that they depicted three scales with exemplars, but they could not distinguish the arrows indicating the values.

Prime trials displayed a weak and a foil picture; targets displayed a weak picture and a ‘blur’ picture. Weak pictures always involved a mismatch between the values in the scales: the two first scales returned low values for the individual tokens (i.e. located at the green portion of the scale), the third scale returned a high value for the two tokens together (i.e. located at the red portion of the scale). Note that this is the case for both distributive and collective primes and targets. Foil pictures made the sentence false by making the predicate false in all three scales.

Besides experimental trials, we also included four types of **control trials** (e.g. Figure 2). *Foil* controls involved an unambiguous sentence (e.g. ‘The bag is heavy’) together with a picture that made the sentence false and a ‘blur’ picture. Participants were thus forced to select the ‘blur’ option. *True* controls were the counterpart of foil cases. They also involved an unambiguous sentence but they displayed one picture that made the sentence true and a ‘blur’, leading participants to choose the overt picture. These controls were designed to highlight both ‘blur’ and overt pictures as possible correct responses, preventing participants to develop verification strategies based on the type of picture.

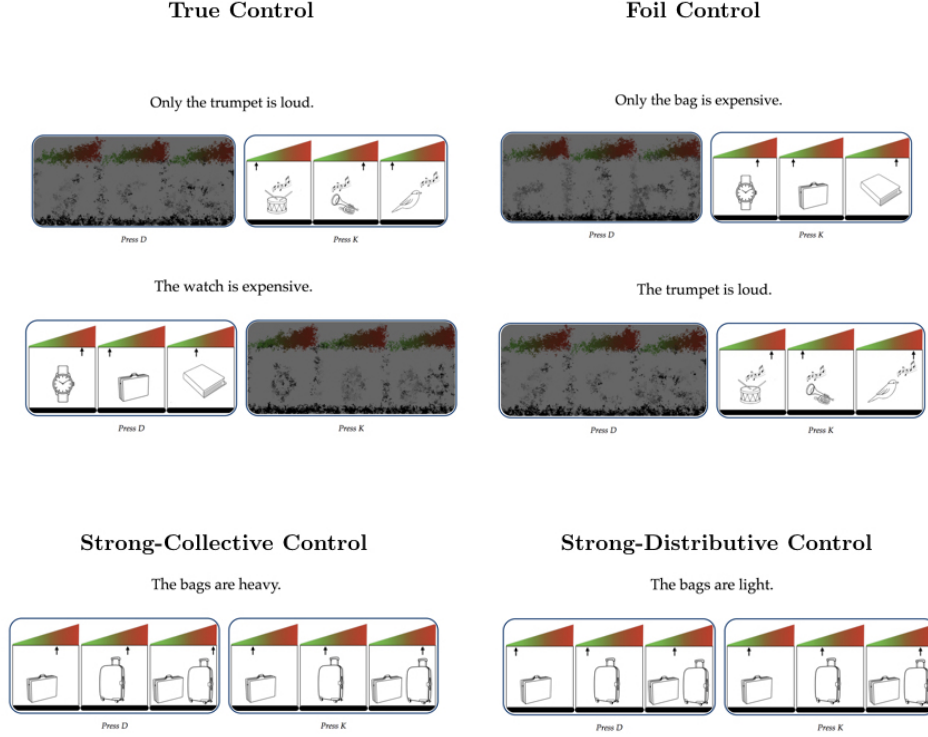
In addition, we included *Strong-Distributive* and *Strong-Collective* controls. These involved the same ambiguous sentences as in primes and targets (e.g. ‘The bags are heavy’), but displayed a weak picture and a *strong* picture, which made both readings of the sentence true. The idea behind these controls was to make participants noticing that the ‘blur’ picture in targets could hide an scenario than makes both readings true (strong picture). In the same way prime trials raise the likelihood of the ‘blur’ option being a foil picture, these *strong* controls raise the likelihood of the ‘blur’ picture being a situation that makes both reading true. On top of elevating the overall proportion of ‘blur’ responses in targets, these controls should give us the baseline preference pattern between strong and weak readings.

### 1.1.3 Design

There were two PRIME CONDITIONS (collective and distributive) and two TARGET CONDITIONS (collective and distributive), giving rise to four possible prime-target combinations. Collective primes and targets always involve positive adjectives, whereas distributive primes and targets involve negative ones (see 1).

In order to boost the priming effect, experimental units were organised in triplets: there were two primes of the same condition preceding each target. Two consecutive primes always involve the same predicate but they differ in the exemplar used in the sentence as well as in the position on the screen for the correct image (i.e. the weak picture)

Primes and target within one experimental unit could instantiate predicates from the same dimension or from different ones (matching or mismatching dimension, PREDICATE CONDITION).



**Figure 2:** Illustration control trials

Moreover, the position of the weak picture in targets could either be that of the weak (correct) image in the first prime or in the second prime (WEAK IMAGE POSITION).

The experimental design consisted on four fully crossed factors to obtain 16 experimental units: 2 (PRIME CONDITION)  $\times$  2 (TARGET CONDITION)  $\times$  2 (PREDICATE CONDITION)  $\times$  2 (WEAK IMAGE POSITION). The total number of experimental trials was 48. There were a further 16 trials for each control condition. These 64 controls trials were randomly inserted between experimental units.

#### 1.1.4 Procedure

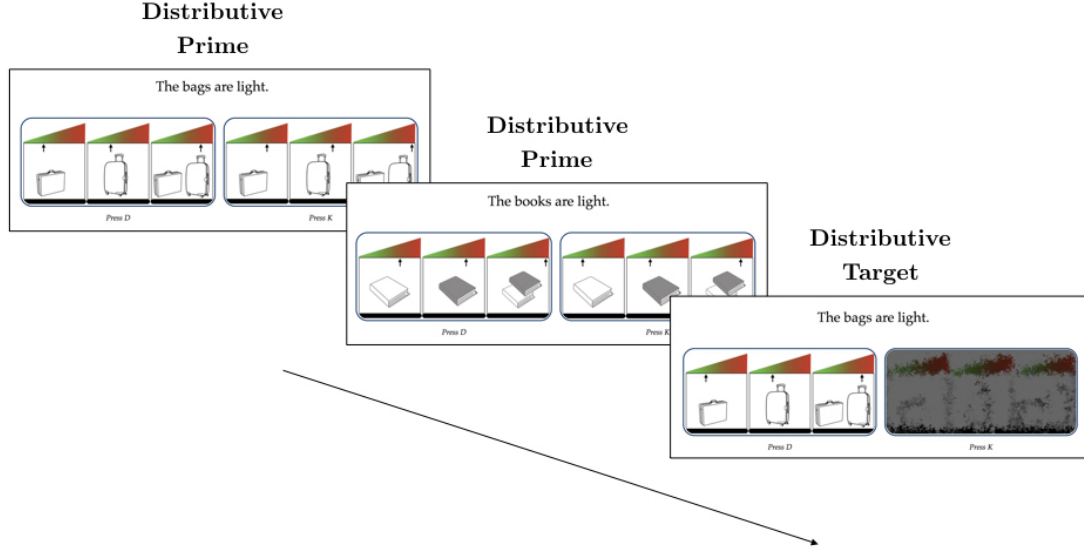
Participants were instructed to select between two pictures the one that best illustrates the sentence. They were given two examples, both involving unambiguous sentences such as the ones used in control trials. One of examples displayed two overt images, and one involved a ‘blur’ option. They were instructed to select the blur picture only if the alternative could not be a good sentence description.

The experiment was implemented using the Ibex Farm online platform. Experimental triplets and controls were randomly administered to each participant. The presentation paradigm is exemplified in Figure 3.

## 1.2 Results

### 1.2.1 Data treatment

Participants with accuracy rates below 75% in True and Foil Controls were removed from the analyses. The remaining 33 participants were taken into account.



**Figure 3:** Illustration of procedure for a Distributive-Distributive triplet

Target trials were preceded by two prime trials. In order to ensure that participants had derived the right interpretation for prime sentences, we removed from the analysis all target responses that were not preceded by two correct prime responses (corresponding to 10% of the targets). General accuracy in prime trials was above 90 % (Collective primes: 97% (SE:.7); Distributive primes: 92%(SE: 2)) .

### 1.2.2 Analysis procedure

Target responses were analysed by modeling response-type likelihood using logit mixed-effect models (CITE). Due to lack of convergence, the random structure included random intercepts per subjects. Analyses were conducted using the lme4 (CITE) library for the R statistics program (CITE).

The p-values were obtained by a  $\chi^2$  likelihood ratio test comparing our model with a simpler one in which the relevant predictor was removed. This type of analysis has been previously used to test similar priming effects (Chemla and Bott, 2015; Raffray and Pickering, 2010, among others).

The dependent measure was the log odds of choosing a distributive over a collective response on target trials. Following Table X,

### 1.2.3 Analysis

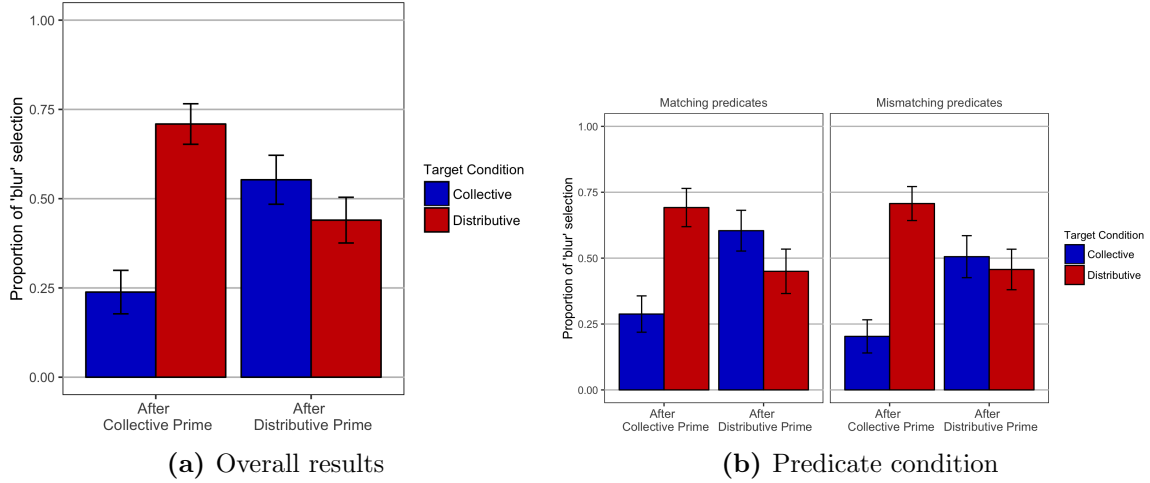
The mean percentage of ‘blur’ responses after accurate primes is illustrated in Figure 4a (i.e. the proportion of target trials in which the ‘blur’ picture was selected per condition). Figure 4b breaks this down depending on whether or not primes and targets had a matching or a mismatching dimension.

We report three analyses. The initial analysis assessed whether semantic priming was at play (i.e. whether primes from different conditions had a different impact on targets). This model took the target condition (levels: collective, distributive), the prime condition (levels: collective, distributive) and their interaction as predictors. The comparison with a simpler model which drops the interaction from the predictors suggest a significant interaction  $\text{PRIME} \times \text{TARGET}$

conditions ( $\chi^2 = 48.9$ ;  $p < .001$ ). After a collective prime people select more often pictures compatible with the collective reading; while after a distributive prime they select more often pictures that are compatible with the distributive reading.

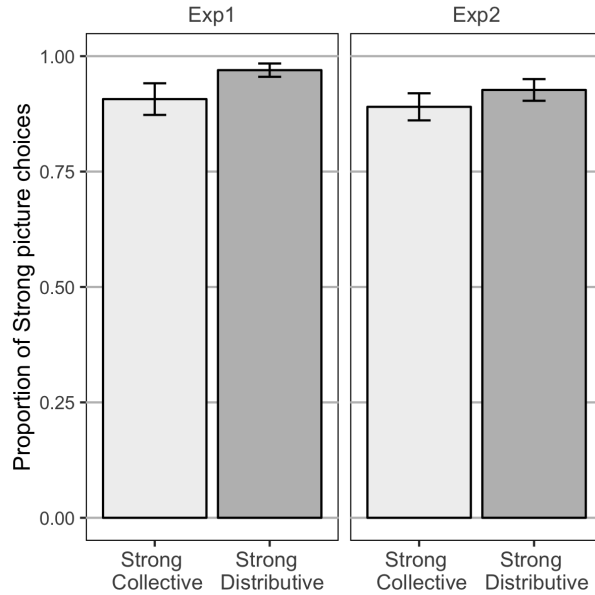
The second analysis No significant interaction with PREDICATE CONDITION interaction (PRIME  $\times$  TARGET  $\times$  PREDICATE conditions:  $\chi^2 < 1$ ;  $p = .81$ ): Both matching and mismatching predicates pairs within one experimental pair give rise to similar effects.

Lastly, the third analysis evaluated whether the influence of one type of prime differ from the



**Figure 4:** Target results in Experiment 1

Responses in Strong controls: Overall preference for pictures that make both readings of the sentence true (*strong* picture). Illustration given in Figure 5.



**Figure 5:** Strong Control results in Experiments 1 and 2

### 1.3 Discussion

Our results suggest that semantic priming is at play: participants' choices in targets are influenced by the reading that was forced in prime trials.

- The presence of an interaction between PRIME and TARGET conditions suggests that we are not facing a simple visual priming effect. If a visual priming effect would be at play, we would expect participants to choose the same kind of images in primes and targets across the board. However, the selection of the overt image in primes (weak picture) biased the selection of the same kind of overt picture *only* when prime and target belong to the same condition (distributive or collective). In contrast, whenever primes and targets were of different conditions, the selection of the overt picture resulted on an increased of blur choices in targets.
- One might still think that visual priming is at play, but participants are sensitive to the polarity of the adjective in such a way that whenever prime and target involve predicates of different polarity (which is equivalent to being of different conditions), they are *anti*-primed: they are primed to *not* choosing the same image in targets as they did in primes. This kind of effect could also give rise to the attested interaction and it's difficult to completely ruled our with our paradigm. However, given that we used different predicates, this hypothesis seems rather unlikely (Subjects should get anti-priming effects for cheap to heavy, for instance). This might be controlled for if we do the across-predicates experiment, from transitive to intransitive and vice versa)

**Asymmetric priming** In Maldonado et al (2017), an asymmetry between distributive and non-distributive priming was found, such that distributive primes could influence target selection, but non-distributive primes behave just like baselines (i.e. targets after no prime).

- Maldonado et al (2017) compared distributive priming to *cumulative* priming. It would be interesting to test whether this asymmetry is also presented for collective vs. distributive.
- Distributive readings are thought to be derived from non-distributive readings, independently of whether they are cumulative or collective. If the specific mechanism responsible for distributivity can be primed, we would expect an asymmetry of priming effect (even when both effects are at play).
- In our experiment, we cannot see whether there is priming in both directions or only in one. In order to see that, we needed a baseline rate of target responses in absence of semantic priming.

**Verification strategies** An alternative explanation of the results could lay on priming of a specific verification strategy.

- When participants face a collective prime, they can decide which picture to choose based on whether the scale containing the 2 tokens satisfies the predicate or not. Instead, when participants face a distributive prime, they need to do exactly the opposite: check for the two left-most scales which have individual tokens.
- After being forced to apply one of these two verification strategies in primes, participants might want to use the same strategy in targets. This would lead to: participants choosing overt pictures more often for Collective-Collective and Distributive-Distributive pairs (when they can apply the strategy), and blur pictures for Collective-Distributive and Distributive-Collective pairs (when they cannot apply the strategy).

- A priming of this verification strategy would result exactly on the same pattern of result we have observed.
- Experiment 2 was designed to rule out this alternative explanation of the results.