



HAL
open science

Understanding semantic ambiguities : an experimental perspective

Mora Maldonado

► **To cite this version:**

Mora Maldonado. Understanding semantic ambiguities : an experimental perspective. Linguistics. Université Paris sciences et lettres, 2018. English. NNT : 2018PSLEE014 . tel-01892104

HAL Id: tel-01892104

<https://theses.hal.science/tel-01892104>

Submitted on 10 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'École normale supérieure

Understanding semantic ambiguities. An experimental perspective.

La compréhension des ambiguïtés sémantiques: une perspective expérimentale.

Ecole doctorale n°158

ED3C - CERVEAU COGNITION COMPORTEMENT

Spécialité SCIENCES COGNITIVES

**Soutenue par Mora Maldonado
le 05 Juin 2018**

Dirigée par **Benjamin Spector
& Emmanuel Chemla**



COMPOSITION DU JURY :

M. SCHLENKER, Philippe
Ecole Normale Supérieure, Président du jury

Mme. WELLWOOD, Alexis
University of Southern California,
Rapporteur

M. SAUERLAND, Uli
Leibniz-Zentrum Allgemeine
Sprachwissenschaft (ZAS), Rapporteur

Mme. SCHMITT, Viola
Université de Vienne, Membre du jury

M. SPECTOR, Benjamin
Ecole Normale Supérieure, Directeur

M. CHEMLA, Emmanuel
Ecole Normale Supérieure, Co-Directeur

Université de recherche Paris Sciences et Lettres
École Normale Supérieure
Département d'Études Cognitives

UNDERSTANDING SEMANTIC AMBIGUITIES
AN EXPERIMENTAL PERSPECTIVE

Mora Maldonado

Supervisors: Benjamin SPECTOR and Emmanuel CHEMLA

Ph.D. Thesis
Institut Jean Nicod
Laboratoire de Sciences Cognitives et Psycholinguistique
École Doctorale 158 Cerveau Cognition Comportement
Defense date: 5th June 2018

Durante un segundo de lucidez tuve la certeza de que nos habíamos vuelto locos. Pero a ese segundo de lucidez se antepuso un supersegundo de superlucidez (si me permiten la expresión) en donde pensé que aquella escena era el resultado lógico de nuestras vidas absurdas.

ROBERTO BOLAÑO, *Los Detectives Salvajes*

Smokey bear: Only who can prevent forest fires?

Bart examines the response panel, which has two buttons, marked 'you' and 'me', and presses 'you'.

Smokey bear: You pressed 'you', referring to me. That is incorrect. The correct answer is 'you'.

THE SIMPSONS, <https://www.youtube.com/watch?v=wX1x7pfH8fw>

Abstract

Sentence ambiguities have been at the center of the research on language comprehension for some time. For semanticists, these ambiguities have been taken to suggest the existence of different abstract mechanisms that may apply to the same syntactic structure at the interpretation stage. For psycholinguists, semantic ambiguities have provided a tool to analyze the dynamics of sentence parsing: since ambiguities tend to be solved incrementally (i.e. before the end of the sentence), the processing pattern of ambiguous sentences might allow identifying the linguistic and non-linguistic factors that play a role during online comprehension.

This dissertation informs theories of language comprehension by exploring two complementary questions: (1) how are different meanings associated to a single sentence form, and (2) how are we able to access and compute these alternative interpretations during parsing. To address these questions, the present work mainly focuses on the so-called plural ambiguities, which arise by the interaction between certain predicates and their plural arguments. For instance, the sentence *Amir and Milica built a sandcastle* has a non-distributive, collective, interpretation (i.e. Amir and Milica *together* built a sandcastle) as well as a distributive one (i.e. Amir and Milica *each* built a sandcastle). Most linguistic approaches assume that distributive readings are derived from more basic non-distributive interpretations by the application of a covert “distributivity” operator (Link, 1983; Champollion, 2014).

The first part of this dissertation presents two studies that aim to identify the abstract mechanisms underlying the distributive/non-distributive contrast through a priming paradigm. This priming method is then extended to other semantic phenomena (i.e. scope ambiguities) in the second part of the dissertation, where some interactions between plurality and scope phenomena are also tested experimentally. To assess the dynamics of ambiguity resolution, the third part of this work presents a mouse-tracking study designed to establish the features of mouse-trajectories that correlate with decision making and disambiguation. The methodology developed in this study is then used to analyse preliminary data on the processing of plural ambiguous sentences.

Resumé

Les ambiguïtés de phrases sont au cœur de la recherche sur la compréhension du langage depuis un certain temps. Pour les sémanticiens, ces ambiguïtés ont été utilisées pour suggérer l'existence de différents mécanismes abstraits qui pourraient s'appliquer à une même structure syntaxique au stade de l'interprétation. Pour les psycholinguistes, les ambiguïtés sémantiques ont offert un outil d'étude de la dynamique du traitement de phrases: puisque les ambiguïtés tendent à être résolues incrémentalement (c'est-à-dire avant la fin de la phrase), le schéma de traitement des phrases ambiguës peut permettre d'identifier les facteurs linguistiques et non linguistiques qui jouent un rôle dans la compréhension online.

Cette dissertation traite des théories de la compréhension du langage en explorant deux questions complémentaires : (1) comment différents sens peuvent-ils être associés à une seule tournure de phrase, et (2) comment sommes-nous capables d'accéder à ces interprétations alternatives et de les traiter pendant l'analyse syntaxique. Pour répondre à ces questions, la présente étude se focalise principalement sur ce qu'on appelle les ambiguïtés de pluriel, qui surgissent par l'interaction entre certains prédicats et leurs arguments pluriels. Par exemple, la phrase *Amir et Milica ont construit un château de sable* a une interprétation non distributive, collective (c'est-à-dire qu'Amir et Milica ont construit *ensemble* un château de sable) mais aussi une interprétation distributive (c'est-à-dire qu'Amir et Milica ont *chacun* construit un château de sable). La plupart des approches linguistiques partent du principe que les lectures distributives dérivent d'interprétations non distributives, plus élémentaires, par l'application d'un opérateur de "distributivité" phonologiquement nul (Link, 1983; Champollion, 2014).

La première partie de cette dissertation présente deux études qui visent à identifier les mécanismes abstraits qui sous-tendent le contraste distributif/non distributif à travers un paradigme d'amorçage. Cette méthode d'amorçage est ensuite étendue à d'autres phénomènes sémantiques (c'est-à-dire des ambiguïtés de portée) dans la deuxième partie de la dissertation, dans laquelle des interactions entre pluralité et phénomène de portée sont aussi testées expérimentalement. Pour évaluer la dynamique de la résolution des ambiguïtés, la troisième partie de cette dissertation présente une étude de suivi des mouvements de souris, conçue pour établir les caractéristiques de trajectoires de souris qui se corrèlent avec la prise de décision et la désambiguïsation. La méthodologie développée dans cette étude est ensuite utilisée pour analyser des données préliminaires relatives au traitement de phrases à ambiguïtés de pluriel.

Acknowledgments

To begin with, I would like to thank the members of my committee: Philippe Schlenker, Alexis Wellwood, Viola Schmitt and Uli Sauerland. I genuinely enjoyed the exchange during my defense with each and everyone of you. Your comments, suggestions and questions have made me rethink some important aspects of my work, refreshing my motivations for it.

When I arrived to Paris, I had a bunch of ideas about what it is to be a scientist, but I had never really thought about what kind of scientist I wanted to be. It is thanks to my supervisors, Benjamin and Emmanuel, that now, almost five years later, I am leaving Paris with a clear idea of who I would like to become. Benjamin and Emmanuel were the perfect supervisor-couple. Their co-supervision had the right balance of pressure, support and care, so synchronised and appropriate that many times I thought it was premeditated. They not only took the time to discuss with me even those ideas that they knew wouldn't work and to correct my unintelligible and ungrammatical writing, but more importantly, they showed me that the best science is done with passion, honesty and generosity. So thank you for all this!

Beyond my supervisors, I also profited from collaborating with many other people, as it becomes clear by looking at the table of contents of this dissertation. Besides playing a fundamental role in the conception and development of the paper on mouse tracking, Ewan Dunbar spent many hours patiently explaining to me different modeling tools and teaching me all I know about them. I have also really enjoyed collaborating with Roman Feiman and Jesse Snedeker, from whom I learned not only how to cleverly control for potential confounds in priming experiments, but also how to challenge some of the assumptions I was generally taking for granted.

I would also like to thank Salvador Mascharenas who, together with Benjamin, gave me the opportunity of TA in his Introduction to Linguistic class and showed me how a wonderful, engaging class is taught. This experience made realize how much I enjoy teaching and gave me the tools I need to do it on my own in the future.

This dissertation would not have been possible without the support of my trench companions: Amir Anvari, Manuel Križ and Milica Denic. In addition to making our office a fun place to be from sunrise to sunset, you have repeatedly interrupted your work to answer each of my questions, teaching me most of the English, semantics and English semantics I know. And, because we are not just office mates but friends, I also need to thank each you: Amir, my not-so-pechito cigarette comrade, for making me think and rethink the important stuff beyond the academic bubble; Manuel, the best possible co-author and dinner-maker, for your immense generosity; Milica, my best friend in these five years, for always having the right hug in the very right moment.

I was very lucky to be part of two different labs at DEC. First, I thank everyone at LSCP for creating such interdisciplinary and friendly environment. Besides learning about topics on cognitive sciences which are beyond the scope of my research, the LSCP meetings and discussions made me be fully aware of the role that scientists have in our community and of the responsibilities that come with it.

I was also part of the *Linguae* group at IJN, which was my academic family. Between talks, Q&As and pizzas (or the fancier sushi alternative), I was not only introduced to the most important topics in Linguistics but also given the opportunity to discuss my own work and obtain essential feedback. Moreover, I had the pleasure of being in co-charge of the IJN Junior Reading Group in Linguistics and Philosophy, where I made an amazing group of PhD friends. Thanks to Philippe and Salvador for supporting this initiative.

There are many people from these two teams who gave me support, feedback and friendship in these past five years: Adriana Guevara Rukóz (my partner in the ED3/ENS bureaucratic

quest!), Julia Carbajal, Alex Martin, Si Berebi, Mireille Babineau, Dan Hoek, Jeremy Kuhn, Rachel Dudley, Brian Buccola, Alex Cremers, Lyn Tieu, Armando Lavalle, and Louis Rouillé. Thank you! I would also like to thank Radhia Accheb, Lucie Ravaud, Vincent Gaudefroy and Nathalie Evin, who have helped me deal with all sort of administrative and technical issues during my PhD.

I also had the pleasure to meet some amazing people in Boston, who made my stay there not only scientifically challenging but also very fun. Thanks Sophie Moracchini, Aurore González, Paul Marty, Fabian Moss, and Kenny Chatain.

Quien ha vivido en una lengua que le es ajena sabe que la magia no está en encontrar con quien poder usar el código conocido, sino con quien decidir no usarlo. Tengo entonces que agradecerle profundamente a mi Telemundo en español (en orden de llegada): a Ignacio Rebollo, mi hermano desde el primer momento en que pisé la ENS, por todos esos domingos familiares; a Hernán Anlló, por haber escuchado mi lamento de portnoy y disuelto mi estrés a lo largo de estos años; a Camila Scaff, por el amor violento que nos deslumbró y que hizo de este último año un gran año; a Andrés Soria Ruiz, por haberme hecho perderle el miedo a los sky-lifts y darme música cuando más la necesité. Que siga la marcha tropical.

Last but not least, me gustaría agradecerle a las tres personas sin las cuales nunca hubiese llegado a estar en donde estoy: a mi papá, por haberme enseñado que tener duende es lo más caro que la vida puede ofrecer no sólo a los intelectuales, sino también a los científicos; a mi mamá, que miró una y otra vez los videos de Pinker con el único fin de entenderme, por haber recibido todas y cada una mis decisiones con alegría; y a Rodrigo, mi oriente próximo, por haber sido mi único héroe en este lío.

Contents

1	Introduction	1
1.1	The sources of sentence ambiguity	1
1.2	A case study: Plural ambiguities	3
1.3	This dissertation: A hopscotch	6
I	Plural ambiguities through priming	9
2	Priming the cumulative/distributive distinction	
	joint work with E. Chemla and B. Spector (2017, <i>JLM</i>)	10
2.1	Introduction: plural ambiguities and priming effects	10
2.1.1	Formal approaches to distributive and cumulative readings	11
2.1.2	Understanding plural ambiguous sentences	13
2.1.3	Goals	14
2.2	Experiment 1	15
2.2.1	Methods	16
2.2.2	Results	18
2.2.3	Discussion	20
2.3	Experiments 2 and 3: Control experiments	24
2.3.1	Methods	25
2.3.2	Results	26
2.3.3	Discussion	27
2.4	General Discussion	28
2.5	Appendix	29
3	Priming the collective/distributive contrast for gradable adjectives	
	joint work with E. Chemla and B. Spector (submitted)	32
3.1	Introduction	32
3.2	Experiment 1	34
3.2.1	Methods and materials	34
3.2.2	Results and discussion	36
3.3	Experiment 2: Baseline inclusion	37
3.3.1	Methods and materials	37
3.3.2	Results and discussion	39
3.4	Conclusions	40
3.5	Supplementary materials	40
4	Plural ambiguities through priming: A brief discussion	43
4.1	A closer look at the non-distributive/distributive contrast	43
4.1.1	Do cumulative and collective interpretations arise from the same representation?	43

4.1.2	Sources of distributivity: transitive predicates <i>versus</i> gradable adjectives	45
4.1.3	Different accounts of the distributive/non-distributive contrast (added on April 23rd)	47
4.2	A note on preference and asymmetry differences	48
4.2.1	Priming the presence of an operator	49
4.2.2	Inverse preference effects relative to baseline preference rates	49
4.2.3	Inverse preference effects relative to prior expectations	50
4.2.4	A hybrid explanation: ceiling effects	51
II	Beyond plural ambiguities	53
5	Priming methods in semantics and pragmatics joint work with B. Spector and E. Chemla (2017, <i>BBS</i>)	54
6	Priming scope ambiguities joint work with R. Feiman (first author) and J. Snedeker (submitted)	57
6.1	Introduction	57
6.2	Experiment 1	61
6.2.1	Methods	61
6.2.2	Results	62
6.2.3	Discussion	63
6.3	Experiment 2	64
6.3.1	Methods	64
6.3.2	Results and Discussion	65
6.4	Summary of Findings	66
6.5	General discussion	67
6.6	Conclusion	69
7	An experimental note on distributivity and scope joint work with M. Križ (in prep.)	70
7.1	Introduction	70
7.1.1	The distributive/non-distributive contrast.	71
7.1.2	Plural quantifiers.	71
7.1.3	Previous psycholinguistic work.	73
7.2	Methods	74
7.2.1	Participants	74
7.2.2	Task and Design	74
7.3	Results	76
7.4	Discussion	77
7.5	Conclusions	79
7.6	Supplement on statistical analysis	79
7.6.1	Analysis 1: Baseline	79
7.6.2	Analysis 2: Distributivity preferences	80
7.6.3	Analysis 3: Scope-taking preferences	82
III	A methodological approach to disambiguation	85
8	Mouse tracking as a window into decision making join work with E. Chemla and E. Dunbar (submitted)	86
8.1	Introduction	86

8.2	Validation Experiment: presentation and qualitative analysis	88
8.2.1	Participants	88
8.2.2	Design	88
8.2.3	Interface	89
8.2.4	Data treatment	89
8.2.5	Overall performance	90
8.3	Validation Experiment: Classifying decision processes with LDA	90
8.3.1	Description of LDA classifier	91
8.3.2	Performance of the LDA classifier	91
8.3.3	Meaningful features and optimal predictors	94
8.4	Validation Experiment: LDA <i>versus</i> traditional mouse tracking analyses	94
8.5	Extension to linguistic data	98
8.5.1	Experiment	98
8.5.2	Results and discussion	100
8.5.3	Other mouse tracking measures	102
8.6	Conclusion	105
9	Mouse tracking techniques applied to plural ambiguities	106
9.1	Introduction	106
9.2	Methods	107
9.2.1	Participants	107
9.2.2	Interface	107
9.2.3	Procedure	107
9.2.4	Materials	107
9.2.5	Design	109
9.3	Results and Discussion	109
9.3.1	Pre-processing of the data	109
9.3.2	Analyses	109
9.3.3	Control results	110
9.3.4	Target results	112
9.4	Conclusions	115
IV	Coming back to plurals: The semantic import of morphological number	117
10	Strong plurality and d-linking in Spanish interrogatives	
	(submitted)	118
10.1	Introduction	118
10.2	Number marking in interrogative sentences	120
10.3	Characterizing <i>quién</i> and <i>quiénes</i> interrogatives	122
10.3.1	Cardinality requirement	122
10.3.2	D-linking requirement	126
10.4	Account	128
10.5	Beyond interrogatives: Conclusions and issues for further research	131

1 | Introduction

Ambiguity is a widespread phenomenon in all levels of linguistic analysis: we can easily find morphemes, words and sentences that have more than one possible meaning out of context. This multiplicity of meanings often goes unnoticed by speakers, and is rarely harmful to actual communication practices, which suggests that we are very good at disambiguating between possible interpretations (Jaeger, 2010; Piantadosi, Tily, & Gibson, 2012).

The ubiquity of ambiguities in natural language has been a challenge for both theoretical and psycholinguistic approaches to language comprehension.¹ Indeed, a theory of language comprehension requires a complete understanding of how multiple meanings can be computed for a single form. This requires not only identifying the elements that trigger the ambiguity, but also explaining how comprehenders access and process these different meanings, rapidly solving the ambiguity.

In this dissertation, I present a collection of chapters that investigates the interpretative processes involved in our understanding of *ambiguous sentences*—cases where one sentence systematically gives rise to more than one interpretation. Specifically, my primary goal is to assess whether the abstract mechanisms proposed by linguists to generate different interpretations have a psychological correlate during comprehension. As a result, this work contributes to linguistic theories by providing psychological evidence that dissociates between conceptually divergent views about how natural languages encode and manipulate ambiguous sentences.

To address this goal, I mainly discuss phenomena surrounding plural ambiguous sentences. This topic has been investigated by formal semanticists at length, giving rise to sophisticated models, which have barely been tested using psycholinguistic means.

1.1 The sources of sentence ambiguity

Ambiguities can have different sources. In what follows, I will distinguish two broad types of ambiguities, depending on whether they are triggered by *structural* or by *non-structural* factors.²

Before entering in the classification, it is important to note that, in this dissertation, I will use the term *ambiguity* to cover cases where the multiplicity of meanings arises in a systematic way, and there is a specific mechanism that can be stipulated to adjudicate between interpretations. That is, I will consider that a sentence is ambiguous as soon as it alternates in truth values, somehow independently of what is the source of this alternation. This understanding differs from the traditional, stronger use of the term. According to Gillon (2004), an expression is ambiguous if and only if it can be assigned more than one truth value with respect to a fixed state of affairs and with respect to a single context of use. This definition of ambiguity encompasses cases where there are different syntactic structures or lexical items that can be mapped into

¹This pervasiveness has also been a central argument in the debate about language evolution. Depending on the view, ambiguities have been seen as a challenging or as a desirable property of communication systems (Chomsky, Belletti, & Rizzi, 2002; Piantadosi et al., 2012; Zipf, 1949). I will not enter into this discussion here.

²I am leaving aside ambiguities arising for sentences that include lexically ambiguous words, such as ‘I saw a bat’. Given that strings like ‘bat’ can be mapped into more than one lexical entry, sentences involving these items will potentially inherit the ambiguity.

the same string (see below), but leaves aside cases where, for instance, the alternation in truth values arises as a function of the context of use. As a result, some of the examples which I will treat here as ambiguities stemming from non-structural factors have been traditionally considered to be cases of context sensitivity or deixis rather than of true ambiguity (Gillon, 2004; Sennet, 2016).

Structural ambiguities arise when a single sentence form can be associated with more than one syntactic representation or Logical Form (henceforth, LF). For clarity, let us define a LF as a syntactic structure that is interpreted by the semantic component (Fox, 2003). A classical example of structurally ambiguous sentences are *attachment ambiguities* such as sentence (1), based on a famous joke by Groucho Marx. This sentence can have two alternative interpretations, depending on who do we think is wearing the pajamas. The two readings differ on whether, in the underlying LF, the phrase ‘wearing pajamas’ is attached to the NP, as in (1a), or to the whole VP, as in (1b).

- (1) I shot an elephant wearing pajamas
- a. [I [_{VP} shot [_{NP} an elephant [wearing pajamas]]]]
 - b. [I [_{VP} shot [_{NP} an elephant] [wearing pajamas]]]

It is easy to note that attachment ambiguities are not dependent on lexical meanings: a similar ambiguity would be obtained even if every word in (1) is replaced by an alternative of the same category (e.g. ‘I saw a man using sunglasses’). Instead, in these cases, two syntactic structures happen to be mapped into the same string during linearisation.

The existence of many LFs that correspond to the same sentence can also be the result of interpretative or syntactic mechanisms that may apply covertly to derive a new LF from a more basic one. For example, the doubly-quantified sentence in (2) is *scopally ambiguous*: it can have two different interpretations depending on which of the two quantifiers (‘every’ and ‘a’) takes scope over the other. In the LF in (2a), the universal takes scope over the existential, matching the order of the quantifiers in the sentence (surface scope interpretation). Instead, in (2b), the scopal relation is reversed between the sentence form and the LF (inverse scope interpretation).

- (2) Every boy kissed a girl.
- a. $\forall x.\text{boy}(x) \rightarrow \exists y.\text{girl}(y) \wedge \text{kiss}(x, y)$
For every boy x , there is a girl y that x kissed y .
 - b. $\exists y.\text{girl}(y) \wedge \forall x.\text{boy}(x) \rightarrow \text{kiss}(x, y)$
There is a single girl y such that every boy x kissed y .

Unlike attachment ambiguities, alternative readings of scopally ambiguous sentences are thought to be derivationally related. That is, most accounts assume that the inverse-scope interpretation in (2b) is derived from the one in (2a) via a covert operation that allows reversing the scope assignment (Fox, 2000; May, 1977, 1985).

Alternatively, the fact that a sentence can have many possible interpretations might have a *non-structural* rather than a structural source. These are cases where a single LF can systematically give rise to different meanings due to contextual sensitivity. Distinguishing between ambiguities triggered by structural and non-structural factors is often a difficult affair: the same phenomenon might be explained by means of mechanisms of different nature. Arguably, a quite uncontroversial example of a non-structural ambiguity concerns sentences such as ‘Mary loves him’. In order to assign truth conditions to the sentence, one needs to assign a reference to the pronoun ‘him’. Since pronoun resolution is determined by the context of utterance, different readings are expected to arise as a function of the context. That is, the sentence has different truth-conditions across contexts. Importantly, this type of ambiguity does not stem from the existence of multiple LFs compatible with the sentence, but just from the inclusion of a context-sensitive expression (or a contextual parameter) in the LF.

1.2 A case study: Plural ambiguities

In this dissertation, I discuss sentence ambiguity by mainly focusing on *plural ambiguous sentences* such as (3). These sentences are typically compatible with collective and distributive scenarios, as illustrated in (3a) and (3b).

- (3) Amir and Milica built a sandcastle.
- a. Amir and Milica jointly built a sandcastle, without each separately doing so.
 - b. Amir and Milica each built a different sandcastle.

The semantics of plural ambiguous sentences have been extensively studied from a theoretical perspective, which has led to a variety of models competing to best account for introspective data. In particular, it has been a matter of some debate whether these different interpretations arise from the existence of two alternative LFs or whether they stem from a single LF, which might be either contextually sensitive or just underspecified. That is, plural ambiguous sentences have been taken to exemplify cases of structural ambiguity, non-structural ambiguity, or plain underspecification. As a result, these ambiguities are a good phenomenon to illustrate how the multiplicity of meanings can be explained by different interpretative mechanisms. Before entering into the details of the different approaches, I review some of the assumptions I will be relying on regarding plurality and associated ambiguities.

Plural entities: the basics Following Link’s mereological approach (Link, 1983, see also Champollion & Krifka, 2014) to plural denotation, plural and singular objects are both *e*-type, differing only on their internal structure. Basic assumptions about object ontology are given in (4). A notational convention, which I follow throughout this dissertation, is the use of \preceq for the relation of parthood, \preceq_{AT} for atomic parthood, and \oplus for sum formation.

- (4) Objects can be atomic or non-atomic, and be related by a parthood relation.
- a. An object a is atomic if it does not have proper parts: if there is no a' such that $a' \prec a$.
 - b. Two objects a and b can be combined by the *summation* operation, where $a \oplus b$ is the least upper bound of a and b (i.e. the ‘smallest’ entity that has a and b as its parts).

The domain of singular and plural entities (D_e) is taken to be closed under sum: if $a, b \in D_e$, then $a \oplus b \in D_e$ (i.e. it has the structure of a *join semilattice*). To account for plural predication, Link proposes the existence of the sum-closure operator in (5), often written as the ***-operator, which can be applied to one-place predicates. The plural is then also closed under sum: if $\llbracket \text{girl} \rrbracket = \{a, b\}$, $\llbracket \text{girls} \rrbracket = \llbracket * \text{girl} \rrbracket = \{a, b, a \oplus b\}$.

- (5) $*X$ is the smallest set such that:
- (i) $*X \supseteq X$ and
 - (ii) $\forall x, y \in *X, x \oplus y \in *X$

Cumulative and distributive entailments Predicates differ with respect to the entailment pattern they display for their arguments. A general property of predicates that can take atomic individuals as arguments is that they give rise to *cumulative entailments*: whenever they apply to two different entities, they also apply to the two entities taken together (Link, 1983).³ For instance, both sentences (7a) and (8a) entail the respective sentences in (7b) and (8b).

³A caveat is in order for *collective* predicates such as ‘form a pyramid’, which cannot take atomic individuals as arguments: **Amir formed a pyramid and Milica formed a pyramid*. Some of these predicates can have cumulative entailments when there is no category mistake: *The girls formed a pyramid and the boys formed a pyramid* \Rightarrow *The boys and the girls formed a pyramid* (see Champollion, 2010; Winter, 2001 for a discussion about different collective predicates and their properties).

- | | | |
|-----|---|-------------------------|
| (6) | a. 'X VP' and 'Y VP' entails 'X and Y VP' | CUMULATIVE ENTAILMENT |
| | b. 'X and Y VP' entails 'X VP' and 'Y VP' | DISTRIBUTIVE ENTAILMENT |

While some of these predicates (e.g. 'smile') always display the reverse *distributive entailment*, others do not ('build'): sentence (7b) entails (7a), but (8a) does not necessarily follow from (8b). This asymmetry in the entailment pattern of (8b) is tightly related to its ambiguous status. As observed in (3), this sentence can have both a collective and a distributive reading. The distributive entailment holds under the distributive reading of the sentence, which we could paraphrase as 'Amir and Milica *each* built a sandcastle'. However, it does not follow under the collective reading, paraphrased as 'Amir and Milica *together* built a sandcastle'. Consequently, the distributive inference can be considered to be *optional* in these cases.

- | | |
|-----|---|
| (7) | a. Amir smiled and Milica smiled. |
| | b. Amir and Milica smiled. |
| (8) | a. Amir built a sandcastle and Milica built a sandcastle. |
| | b. Amir and Milica built a sandcastle. |

When predicates with plural arguments are pluralized using the *-operator, cumulative entailments are guaranteed. For example, one can easily note that the following entailment holds: $\text{smile}(a) \wedge \text{smile}(m) \Rightarrow * \text{smile}(a \oplus m)$. For two-place predicates (e.g. 'built'), the *-operator needs to be generalized to be applied to the set of pairs that compose the predicate extension. This is done by assuming that the *-operator closes the predicate under *pointwise sum* as defined in (9) (Krifka, 1992; Landman, 1996, 2000).

- (9) For any lexical relational predicate P , its closure under pointwise sum $*P$ is the minimal relation such that: $\forall x_1, y_1, x_2, y_2$ if $P(x_1, y_1)$ and $P(x_2, y_2)$, then $*P(x_1 \oplus x_2, y_1 \oplus y_2)$.⁴

Note that, by extending the application of the *-operator to many-place predicates, we can also account for *cumulative* interpretations of sentences involving more than one plural expression. For instance, (10) is compatible with a situation where Amir invited one linguist to the party and Milica invited another linguist to the party. If the extension of 'invite' contains $\langle a, l_1 \rangle$ and $\langle m, l_2 \rangle$, then '*invite' will contain $\langle a \oplus m, l_1 \oplus l_2 \rangle$, verifying this interpretation of (10).

- (10) Amir and Milica invited two linguists to the party.

A common assumption in plural semantics is that lexical predicates are closed under pointwise sum "from the start". This is the so-called *lexical cumulativity hypothesis* (Champollion, 2010; Kratzer, 2007; Krifka, 1992; Scha, 1981, 1984). Under this view, if a predicate is true of two individuals, it is automatically true of the sum of those individuals. In other words, $\llbracket \text{smile} \rrbracket = \llbracket * \text{smile} \rrbracket = \{a, m, a \oplus m\}$. In this dissertation, I will adopt this view and take predicates to be closed under sum, without including the *-operator in the typographical representation of verb meanings (cf. Kratzer, 2007).

'Mixed' predicates How can we account for differences in distributive entailments? As observed for (8), we can have a predicate P such that $P(a \oplus b)$ is true without $P(a)$ being true. It is often proposed (Champollion, 2010; Landman, 1996, 2000; Lasnik, 1989; Link, 1983; Winter, 2001) that the availability of distributive entailments has a lexical basis. Predicates might include specific requirements as part of their lexical entries (as meaning postulates), or directly differ on the types of entities they can have in their original extension. That is, they can refer to properties of singular individuals (distributive predicates), plural individuals (collective

⁴Note that nothing prevents $x_1 = x_2$ and $y_1 = y_2$. As a result, $P(x, y)$ actually entails $*P(x, y)$.

predicates) or both (“mixed” predicates).⁵

For instance, ‘smile’ is considered to be a *lexically distributive* predicate, which refers exclusively to properties of atoms. In contrast, predicates such as ‘built’ are “mixed”: they can contain (pairs of) both atoms and pluralities in its original extension. If the denotation of ‘smile’ only contains atoms, the distributive entailment for “*smile’ is automatically satisfied: for every plurality X and every atomic entity x such that $x \preceq X$, if *smile(X) then *smile(x). Instead, since ‘built’ is “mixed”, its pluralized version “*built’ may not refer to properties of atoms ($\llbracket \text{built} \rrbracket = \llbracket \text{*built} \rrbracket = \{ \langle a \oplus m, s \rangle \}$, where s is a sandcastle). Thus, the distributive entailment is not guaranteed.

Deriving the distributive/non-distributive ambiguity Let me now return to the original question of how to account for the multiple meanings arising for sentences such as (3). A common view is that non-distributive and distributive readings of sentences like (3) correspond to two alternative LFs (Champollion, 2014, to appear; Link, 1987). That is, these are *structurally ambiguous* sentences.

Under the lexical cumulativity hypothesis, the non-distributive reading of (3) would result from simply applying the predicate to the plural subject, as in (11). This LF makes the sentence compatible with non-distributive scenarios such as (3a). In contrast, the distributive reading –compatible with (3b)– is assumed to be derived through the insertion of a covert distributivity operator, which we will call D , at the level of the verb phrase (see (12), Champollion, 2014). This D operator can be thought of as a silent version of floating ‘each’ in English (modulo homogeneity, see Križ, 2015) that introduces universal quantification over the individual members of the subject.

(11) Amir and Milica built a sandcastle.

$\exists y. \text{sandcastle}(y) \wedge \text{built}(a \oplus m, y)$

There is a sandcastle y such that the plurality made up by a and m built y .

(12) $\llbracket D \rrbracket = \lambda P. \lambda x. \forall y. [y \preceq_{AT} x \rightarrow P(y)]$

in words: given a predicate P and a plurality x , $D(P)$ is true of x iff P is true of all the atomic individuals y that make up x .

The LF resulting from applying D to the predicate (13) ensures the distributive entailment, and allows the indefinite to covary with members of the subject, making the sentence compatible with (3b).

(13) Amir and Milica D built a sandcastle.

$\forall x. x \preceq_{AT} a \oplus m \rightarrow \exists y. \text{sandcastle}(y) \wedge \text{built}(x, y)$

For each atomic member x of the plurality made up by a and m , there is a sandcastle y such that x built y .

Distributive readings, however, do not *impose* covariation. For instance, the distributive reading of the non-creation predicate in (14) is compatible with a scenario where Amir and Milica each carried the same box upstairs. Arguably, this scenario also makes true the non-distributive reading of the sentence: under lexical closure under sum, if ‘carry’ contains the set of pairs $\langle a, b \rangle$ and $\langle m, b \rangle$, then it also contains $\langle a \oplus m, b \rangle$ (Champollion, 2010).

(14) Amir and Milica carried a box upstairs.

⁵I am ignoring here alternative accounts, which propose a two-way distinction between *atom* and *set* predicates, according to the predicate behaviour in sentences involving plural and singular quantificational phrases (Vries, 2015; Winter, 2001). For example, while the sentences ‘All the students smiled’ and ‘Every student smiled’ are equivalent, the sentences ‘All the students met’ and “*Every student met’ are not. This suggests that ‘smile’ is an atom predicate, which, when uninflected for number, ranges over atomic entities, whereas ‘meet’ is not, and it ranges over pluralities (or sets) instead.

The view presented so far is that the non-distributive/distributive contrast reflects a structural ambiguity. Alternatively, some approaches have proposed that this distinction arises from different interpretations of a single logical form. Let me briefly mention two such accounts.

[Schwarzschild \(1996\)](#) offers an analysis of the distributive/non-distributive distinction in terms of a non-structural ambiguity, analogous to the one observed for pronouns. In a nutshell, [Schwarzschild](#) proposes a modified distributivity operator, which does no longer quantify only over atoms, but can instead distribute over non-atomic parts of a salient plurality partition (known as a *cover*)⁶. To interpret this distributivity operator, one needs to determine the way in which pluralities are partitioned, namely, the cover. Crucially, the cover, like pronouns, is contextually determined (i.e. it is a contextual parameter). The sentence truth-conditions will then depend on what kind of cover is pragmatically given. For instance, the collective reading of (3) would result from a relevant cover that offers a single cell containing the plurality $a \oplus m$. Conversely, a cover that divides the plurality into atomic individuals will result in the distributive reading. Interestingly, this account can generate many intermediate readings, which are not always introspectively available.

Alternatively, [Kratzer \(2007\)](#) puts forward an underspecification account of the distributive/non-distributive contrast. First, Kratzer closes the predicate ‘build’ under sum, allowing the sentence to be compatible with all non-covariation scenarios as described above. Then, a second application of the *-operator has as a consequence that the predicate ‘build a sandcastle’ can be applied not only to sums of atoms, but also to atoms, depending on how the plurality is divided into *subparts*.⁷ As a result, by starring the sister of the plural subject in (15a), Kratzer derives the underspecified reading in (15b), whose truth-conditions are compatible with different types of situations (non-distributive and distributive ones).⁸

- (15) a. $[a \oplus m]_1 *(\lambda 1. [\exists y. \text{sandcastle}(y) \wedge \text{built}(t1, y)])$
 b. The plurality made up by a and m can be divided into a set of subparts such that each subpart built a sandcastle.
 \rightsquigarrow comes out true if Amir and Milica each built a sandcastle.
 \rightsquigarrow comes out true if Amir and Milica built a sandcastle together.

While introspective evidence is sometimes hard to produce and evaluate to adjudicate between these positions, preliminary psycholinguistic studies have argued so far for a genuine structural ambiguity. For example, in an eye-tracking study, [Frazier, Pacht, and Rayner \(1999\)](#) observed that people commit early on to one particular reading of sentences such as (14). Under the assumption that people tend to make the minimal semantic commitment whenever they can (i.e., whenever the LF can be left underspecified during semantic computations), the authors conclude that this is a case of a true structural ambiguity, where selecting among two readings is a necessary choice because there are two competing syntactic representations or LFs.

1.3 This dissertation: A hopscotch

I propose below one possible way to navigate this dissertation. The reader, however, is free to choose her own path. For any random chapter the reader chooses, she will be able not only to read it and understand it without any additional background, but also to find a second chapter that relates to it.

⁶Relying on a set-based representation of plural individuals (not mereological, as I have assumed), [Schwarzschild](#) defines a cover as a subset of a set where cells can overlap.

⁷Following the extensional mereology account, a subpart can be thought as any portion of a given entity (atomic or not), even the entity itself: $a \prec a$ ([Champollion & Krifka, 2014](#)).

⁸Note that there is good evidence that the reading obtained by applying the *-operator to the predicate ‘build a sandcastle’ might lead to the wrong results, making the sentence true in scenarios where it is intuitively false. For instance, it would make the sentence ‘Amir, Milica and Manuel built a sandcastle’ true when Amir built a sandcastle on his own, and Milica and Manuel built one together.

This dissertation is divided in four parts. The first part, which comprises Chapters 2 to 4, makes use of a priming paradigm to explore the abstract interpretative mechanisms at play in our understanding of plural ambiguous sentences. In the second part (Chapters 5 to 7), I go beyond plural ambiguities, showing that similar questions and methodology can be applied to investigate phenomena surrounding scope ambiguous sentences. The third part (Chapters 8 and 9) is concerned with the methodological challenge of investigating semantic processing and ambiguity resolution during online comprehension. In the fourth and last part of the dissertation (Chapter 10), I come back to issues related to plurality, and provide a theoretical contribution to the debate about the denotation of plural morphology. In what follows, I give a short outline of these various chapters.

CHAPTER 2 presents three experiments showing that the distributive/cumulative contrast for sentences such as ‘Two boys have three balloons’ can give rise to priming effects. When subjects are forced to access one specific reading in prime trials, they are more likely to access the same interpretation in a subsequent target. This priming effect is stronger for distributive than for cumulative readings, suggesting that abstract constructs such as the distributivity operator describe real features of mental representations.

CHAPTER 3 reports the results of two priming experiments revealing that distributive and collective readings of sentences involving gradable adjectives (e.g. ‘The bags are heavy’) can also be specifically primed. These findings indicate that there is a substantial difference in the derivation of distributive and collective interpretations, which is independent of considerations of object covariation, logical strength and verification strategies.

CHAPTER 4 provides an extended discussion of the results reported in Chapters 2 and 3. I start by exploring in detail the semantic contrasts tested in these experiments: first, I analyze how potential differences between cumulative and collective readings could have affected the results; second, I discuss whether or not the same derivational mechanisms underlie distributive readings of transitive and adjectival predicates. Finally, I explore the differences in priming strength visible in the results of the two studies. Specifically, only one of the studies provides evidence for stronger distributive than non-distributive priming. I comment on this puzzle, providing additional suggestions to understand this data.

CHAPTER 5 presents a short commentary that discusses how a priming method can be used to inform semantic theories. I claim that, while traditional methods in experimental semantics (i.e. acceptability and inferential judgments) serve to reveal which readings arise for a given sentence, priming can target the elementary operations themselves, providing a look into what the sentence is made of.

CHAPTER 6 addresses the question of whether or not we can prime the mechanism required to derive inverse-scope interpretations of scopally ambiguous sentences. I present two experiments revealing that prior evidence for the absence of priming of this operation (see [Chemla & Bott, 2015](#)) needs to be reinterpreted as priming related to pictures and specific spatial verification strategies. Consequently, the possibility of priming the scope-inversion operation remains open.

CHAPTER 7 reports an experiment that tests the existence of distributive readings with plural noun phrases in object position (e.g. ‘A mouse is painting all the penguins’), which has been the subject of some disagreement in the literature. The results of this experiment indicate that, despite being quite marginal, inverse-distributive readings exist for a range of plural quantifiers. Moreover, the availability of these readings is shown to be partly conditioned by quantifier

preferences regarding both distributivity and scope.

CHAPTER 8 investigates the mouse-tracking methodology by asking directly how decision processes translate into mouse trajectories. I present two experiments: one in which the stimuli are explicitly manipulated to trigger a flip in decision, and one in which I replicate more ecological, classical mouse tracking results on linguistic negation (Dale & Duran, 2011). The conclusion of these experiments is that spatial information (mouse path) is more important than temporal information (speed and acceleration) for detecting decision changes. This is done by using an ‘optimal’ analysis of our data: a linear discriminant analysis explicitly trained to classify trajectories.

CHAPTER 9 uses a mouse tracking paradigm to investigate how different interpretations of plural ambiguous sentences are derived during online comprehension. Under the assumption that decision making and disambiguation have some core properties in common, the methodology developed in Chapter 8 is used in this chapter to analyze mouse-tracking data obtained from a sentence-verification task involving plural ambiguities.

CHAPTER 10 presents a study on the semantic import of plural marking in Spanish interrogative words. I analyse Spanish bare interrogative *quién* (‘who_{sg}’) and its plural counterpart *quiénes* (‘who_{pl}’). I argue that the distribution of these interrogatives in Spanish can be accounted for by assuming that the plural *quiénes* triggers a strong plurality presupposition, and can only be used in *d-linking* contexts, whereas *quién* carries no specific requirement, as far as its semantics is concerned. My proposal reveals that current approaches to number marking need to be refined in order to account for cross-linguistic and within-language variation.

Part I

Plural ambiguities through priming

Are the abstract rules proposed to explain the non-distributive/distributive contrast accessed as such during comprehension? In Chapters 2 and 3, I present two studies that use a priming method to address this question. The results of these experiments reveal that these abstract semantic mechanisms are indeed at play during the interpretative process and can be locus of priming. Our findings are further discussed in Chapter 4.

2 | Priming the cumulative/distributive distinction*

Maldonado, M., Chemla, E. & Spector, B. *Priming plural ambiguities*. *Journal of Memory and Language*. 2017.

Abstract

Sentences that involve two or more plural expressions, such as numerical expressions, give rise to systematic ambiguities. For example, the sentence *Two boys have three balloons* can either mean that there are two boys who, between them, have three balloons (*cumulative reading*) or that there are two boys who each have three balloons (*distributive reading*). In this paper, we report the results of three experiments which show that the distributive/cumulative ambiguity can give rise to *priming effects*. That is, when subjects perform a sentence-picture matching task which creates a strong bias towards one of the two types of readings, they are more likely to access the very same type of interpretation when subsequently presented with a different sentence-picture pair which does not create the same bias. This finding suggests that the abstract constructs that linguists posit to account for different types of readings describe some real features of mental representations.

2.1 Introduction: plural ambiguities and priming effects

Sentences that involve two or more plural expressions, such as numerical expressions, can give rise to multiple readings. For example, a sentence like (1) has at least two different readings, paraphrased in (1a) and (1b). Situations illustrating each reading are depicted in Fig. 2.1.

- (1) Two boys have three balloons.
 - a. **Cumulative reading:** There are two boys who, between them, have three balloons (Fig. 2.1a).
 - b. **Distributive reading:** Two boys have *each* three balloons (Fig. 2.1b).

The multiple readings to which plural expressions give rise in natural languages have been extensively studied in theoretical linguistics. Each of the two interpretations mentioned above for this sentence are assumed to instantiate different types of interpretation rules and possibly different syntactic structures, which play a role in many other sentences and sentence types. These rules and structures are, therefore, of an *abstract and general* nature, and it could be assumed that, when processing such sentences, interpreters access these abstract properties of sentences. The general goal of this paper is to investigate whether abstract semantic properties

*We wish to thank Sam Alxatib, Seth Cable, Alexandre Cremers, Isabelle Dautriche, Christophe Pallier, as well as the audiences at NELS 45, Linguae (Institut Jean-Nicod) and Integrative Neuroscience Laboratory (University of Buenos Aires). This paper has benefited from the suggestions and comments of three anonymous *JML* reviewers. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.313610 and from the Agence Nationale de la Recherche (Grants ANR-10-LABX-0087 IEC, ANR-10-IDEX-0001-02 PSL* and ANR-14-CE30-0010-01 TriLogMean)

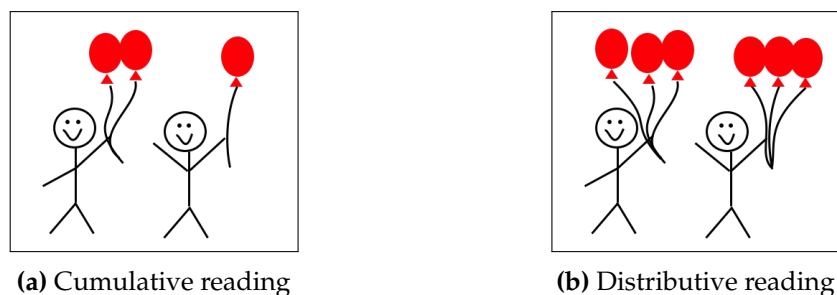


Figure 2.1: Readings of sentences with plurals

of these kinds of plural sentence are accessed during sentence comprehension, by means of a *priming paradigm*.

Before providing more details about our goals and methodology, we introduce some formal semantics background regarding the cumulative/distributive ambiguity and present some of the previous psycholinguistic literature on the processing of plural expressions, ambiguities and priming effects.

2.1.1 Formal approaches to distributive and cumulative readings

Formal semantics approaches to the cumulative vs. distributive contrast are based on the idea that the very same sentence (viewed as a phonological string) corresponds to (at least) two distinct logical forms. A common approach consists in assuming that the distributive interpretation is derived through the application of a so-called *distributivity operator*. More specifically, under the *lexical cumulativity hypothesis* (Krifka 1992; Landman 2000, see also Kratzer 2007 for relevant discussions), the cumulative reading is in some sense primitive, and an operator can be optionally introduced in order to generate the distributive reading. Let us be a bit more specific. In a sentence such as (1), we start with a denotation for ‘have’ which is a relation between *individuals* (called ‘atomic individuals’ in formal semantics approaches to plurality, cf. Link 1983). But when *have* is used in a sentence, it always receives a more complex meaning, which allows it to denote a relation between *sets of individuals*, also called *pluralities* (this is the aforementioned *lexical cumulativity hypothesis*). This relation, which we can represent by HAVE, is defined in terms of the more basic relation ‘have’ as follows. Given two pluralities X and Y , the relation $HAVE(X, Y)$ holds just in case there are pairs (x, y) , where x and y are atomic individuals included in X and Y , respectively, such that x and y are in the primitive relation ‘have’, and there are enough such pairs so that every atomic member of X and every atomic member of Y occurs in at least one such pair (i.e. all of X and all of Y should be ‘involved’ in the ‘have’ relation). This is illustrated in (2), using the verb *own* instead of *have* (because using ‘have’ would degrade the sentence, due to the so-called ‘definiteness effect’, cf. Milsark 1974):

- (2) These boys own these balloons.
 Meaning: everyone of these boys owns at least one of these balloons, and everyone of these balloons is owned by at least one of these boys.

If the referential expressions ‘these boys’ and ‘these balloons’ are replaced with indefinite numerical phrases, the resulting sentence inherits the cumulative reading of HAVE, and as a result (1) is interpreted as follows:

- (3) Cumulative Interpretation of (1):
 There is a plurality made up of two boys, call it X , and a plurality made up of three balloons, call it Y , such that every member of X owns at least one member of Y , and every member of Y is owned by at least one member of X .

It is important to note that the cumulative interpretation of (1) can be true even if no boy individually owns three balloons.

In contrast with this, the distributive reading of (1) entails that there are two boys who each have three balloons. Following Link (1983) or Champollion (to appear), we assume that this second reading is obtained by applying a distributivity operator, noted Δ , to the predicate *have three balloons*. Under this view, while *have three balloons* is true of a plurality X if the members of X have between them three balloons *in total*, $\Delta(\textit{have three balloons})$ is true of a plurality X only if each atomic member x of X has three balloons in total. When the subject of *have three balloons* is an indefinite numerical phrase such as *two boys*, the resulting meaning is that there are two boys such that each of them has three balloons. To sum up, the cumulative and distributive interpretations of (1) correspond to two distinct Logical Forms, which we give in (4) below in a simplified form¹:

- (4) Two boys have three balloons
- a. **LF for the cumulative reading:** [Three boys] [have [three balloons]]
 - b. **LF for the distributive reading:** [Three boys] [Δ [have [three balloons]]]

At this point, we need to note a complication. Under the cumulative reading as defined here, the cumulative interpretation of (1) is true not only in cumulative situations of the type represented in Figure 2.1a, but also in ‘distributive’ situations of the type represented in Figure 2.1b. This is so because in this picture it is possible to single out a plurality of exactly three balloons such that the two boys jointly have these three balloons (take for instance two balloons held by the boy on the left and one balloon held by the boy on the right). In fact, on our assumptions so far, any situation that makes the distributive reading true also makes the cumulative reading true. Indeed, in any such ‘distributive’ scenario, there are 6 balloons that the boys have in total between them, and it is always possible to find a group of 3 balloons that they have between them. However, the cumulative reading of sentences such as (1), which can be forced by adding the phrase *between them* (‘Two boys own three balloons between them’), tends to be strengthened with a pragmatic inference, whose content is that the two boys who have three balloons between them do not have *more* than three balloons between them (see, among others, Landman 2000 for a discussion of such effects produced by numerals in cumulative sentences). That is, (1) can easily be interpreted with an *exact* meaning, i.e. as conveying that there are two boys who between them have *exactly* three balloons. Under this ‘strengthened’ cumulative reading, (1) is false in the type of situations represented in Figure 2.1b. In most of this paper, when we talk about the ‘cumulative’ interpretation, we intend to refer to this strengthened cumulative interpretation. But we will discuss the *exact/at least* contrast and its potential implications regarding the interpretation of our results in section 2.2.3.

¹This presentation does not justice to the complexities of the theoretical literature on plural semantics. In particular, many authors assume, on top or instead of the distributivity operator, that another operator, known as the ‘star’ operator, can apply to one-place predicates and give rise to truth-conditions which are neither equivalent to the cumulative reading nor to what we call here the distributive reading. For (1), the resulting reading is a reading that is very weak, and is entailed by both the cumulative reading and the distributive reading. Under this reading, the sentence would mean that there is a plurality of two boys which can be divided into subparts (not necessarily atomic subparts) such that each subpart has three balloons. In a ‘cumulative’ situation where there are two boys who jointly have three balloons between them, this reading is true because we can always use the trivial ‘division’ of a plurality into subparts where the plurality is ‘divided’ into just one subpart (namely, itself). In a ‘distributive’ situation, the sentence is true as well by considering the division of two boys into two atomic subparts. Kratzer (2007) even assumes that there is no distributive reading *per se*, but only a cumulative reading and this very weak reading that results from applying the star-operator. Champollion (to appear), on the other hand, assumes the star-operator is restricted to apply to lexical predicates, and assumes that there is a distributivity operator, defined as above, which can apply to all one-place predicates. Here, we follow Champollion (2014, to appear).

2.1.2 Understanding plural ambiguous sentences

In English, plurality can be expressed through a number of means, including the use of plural definites (“the boys”), numerical expressions (“two boys”) and quantificational phrases (“each boy”). Many psycholinguistic approaches to plurality have attempted to define how these different plural expressions are interpreted (Kaup, Kelter, & Habel, 2002; Patson, 2014; Patson, George, & Warren, 2014; Patson & Warren, 2010; Sauerland, Anderssen, & Yatsushiro, 2005).

In a recent study, Patson, George, and Warren (2014) presented participants with a sentence together with a picture, and asked them to judge whether the elements named in the sentence appeared in the image. When the sentences contained numeric expressions (“two boys”), people were faster at making judgments about multiple than about singular referents in the picture. In contrast, no differences in time were found for plural definite descriptions. The authors interpret these results as evidence for the existence of different levels of plural representation: while plural definites are conceptualized in an “underspecified” way (i.e. the numerosity remains unresolved), numerically quantified expressions are interpreted as true pluralities. Specifically, numerals appear to be interpreted as referring to *exact* quantities by both adults and children (Huang, Spelke, & Snedeker, 2013; Marty, Chemla, & Spector, 2013; Patson, George, & Warren, 2014): ‘two boys’ would refer to a plurality composed by *exactly* two boys. However, this does not necessarily imply that ‘two’-quantified plurals cannot refer to *more than two* entities: weaker, *at least*, readings of numeric expressions, such as the ones proposed by certain accounts of bare numerals, are considered to be available during parsing (L. Bott & Chemla, 2016). A more extensive theory of how different plural expressions are conceptualized during parsing is provided in Patson (2014).

The study of how these plural expressions interact with each other, giving rise to ambiguities, has been focused on determining whether alternative interpretations (distributive, collective, and cumulative) differ in terms of preference or cognitive cost. Early on, it was observed that adults often prefer collective interpretations of plural ambiguous sentences. Several studies have shown that this pattern is consistent across sentences containing different plural expressions, ranging from numerically quantified phrases (e.g., as in example (1)) and plural definites, to personal pronouns and coordinate noun phrases (Brooks & Braine, 1996; Dotlacil, 2010; Kaup, Kelter, & Habel, 2002; Musolino, 2009; Syrett & Musolino, 2013; Ussery, 1998). This evidence suggests that interpretative differences among plural expressions (such as the ones described above, Patson, 2014) directly favor particular interpretations of ambiguous sentences.

Furthermore, adult preference seems to correlate with the cost associated with each reading (i.e. dispreferred readings are also slower or more costly). For instance, Frazier et al. (1999) used temporally ambiguous sentences (e.g. “Mary and Paul won 100 dollars each/together”) to investigate the dynamics of ambiguity resolution during online processing (see also Brasoveanu & Dotlačil, 2015). These experiments aimed to determine whether people make semantic commitments before disambiguation (e.g., location of ‘each’/‘together’). The authors found a slowdown in reading times (similar to garden-path effects) when sentences were disambiguated towards a distributive reading (but not towards a collective interpretation). These results suggest an early collective interpretation of the predicate, which may be guiding the slowdown in distributive sentences.

Little attention has been paid in the psycholinguistic literature to the mechanisms that allow the derivation of each of these possible interpretations to begin with. In this paper, we will tackle this issue with a *priming paradigm*. Our main goal is to seek behavioural evidence in favor or against the possibility that very abstract constructs of the kind that theoretical linguists posit are active in the actual process of interpreting plural ambiguous sentences.

Priming studies

Psychologists have demonstrated that the processing of one stimulus can be influenced by previous exposure to similar stimuli (Meyer & Schvaneveldt, 1971). This influence, known as a *priming effect*, can reflect the processes that underlie human behaviour, since, for a priming effect to exist, the relevant representations, operations or features shared between the stimuli would have to remain psychologically active across time.

In the current study, we will present a structural priming paradigm. Structural priming effects occur when the activation of one linguistic structure influences the processing of the same structure on a subsequent trial (Bock, 1986; see review in Pickering and Ferreira 2008). These effects have been traditionally attested in language production, across different syntactic structures (e.g., passive/active sentences, double object constructions, among others; Pickering & Ferreira, 2008). Many of these studies have shown that syntactic priming interacts with preference patterns, giving rise to *inverse-preference effects*: less preferred or less frequent constructions exhibit stronger priming effects than more preferred ones compared to a neutral baseline (Bock & Loebell, 1990; Ferreira & Bock, 2006; Hartsuiker & Kolk, 1998; Scheepers, 2003, among others). For example, while English active sentences do not seem to give rise to priming, less frequent passive structures are known to produce strong priming effects (Bock, 1986). This inverse-preference effect has been interpreted as reflecting that priming corresponds to an attempt to minimize prediction error (Chang, Dell, & Bock, 2006; Fine, Jaeger, Farmer, & Qian, 2013). Since dispreferred linguistic structures are less expected in the upcoming discourse, they give rise to a larger prediction error when they do arise. The predicted probability of these infrequent structures will then be increased more strongly for subsequent sentences.

More recently, structural priming has also been shown to facilitate sentence comprehension across different syntactic structures (Thothathiri & Snedeker, 2008), and to influence the interpretation of scopally ambiguous sentences (Chemla & Bott, 2015; Feiman & Snedeker, 2016; Raffray & Pickering, 2010). Similar effects were obtained for semantico-pragmatic phenomena, such as scalar implicatures, where enrichment mechanisms are modulated through priming (L. Bott & Chemla, 2016). No priming study, to our knowledge, has addressed purely semantic contrasts such as the cumulative/distributive ambiguity in comprehension. Our experiments might thus serve to enlarge the range of known linguistic phenomena that display priming effects.

2.1.3 Goals

Our main goal in this paper is to determine whether the cumulative–distributive contrast can give rise to *priming effects*. The underlying question is whether the derivation of, for example, a distributive reading for one sentence at one point in time, biases the interpretation of a subsequent *different* sentence, by increasing the likelihood of a distributive interpretation for the second sentence. Such a finding would confirm that certain *abstract semantic properties*, such as the cumulative–distributive distinction, are fully distinguishable during parsing and can be accessed in certain types of decision tasks.

A secondary goal is the following: if a priming effect is found, we would like to determine whether such a priming effect is *symmetric* or *asymmetric* with respect to the cumulative/distributive contrast. Here, we define the priming effect as *symmetric* if the bias towards the distributive interpretation when subjects are primed with a distributive interpretation and the bias towards the cumulative interpretation when subjects are primed with a cumulative interpretation are equally strong.

According to (4), the cumulative reading is in some sense primitive, and the LF for the distributive reading is obtained from the LF for the cumulative reading by adding the distributivity operator. If the priming effects found (if any) are at least partly triggered by the distributivity operator, we may expect stronger priming effects with distributive primes than with cumulative

primes (*asymmetric priming*).

Furthermore, an asymmetric priming could suggest an inverse-preference effect, such as the one described above for syntactic priming. The preferred reading of plural ambiguous sentences could demonstrate less priming with respect to a baseline than the dispreferred one. This possibility could still be related to the LFs in (4): the fact that distributive interpretations might involve more complex representations (i.e. with an extra operator) could affect the preference and, consequently, the strength of the effect.

In the rest of this paper, we present the results of three priming experiments which shed light on these questions.

2.2 Experiment 1

We used a sentence-picture matching task (Raffray & Pickering, 2010) where participants had to match a sentence with one of two pictures. The sentences always refer to relations between shapes, but they could be ambiguous (e.g., “Two circles are connected to three squares”) or unambiguous (e.g., “A circle is connected to three squares”), depending on the trial. Experimental trials were *primes*, *baselines* or *targets*. Examples are shown in Fig. 2.2.

In prime trials, the sentence was ambiguous between a *cumulative* and a *distributive* interpretation (e.g., “Two circles are connected to three squares”). One of the two images corresponded to one of these readings (henceforth called the ‘correct’ picture), while the other was incompatible with both readings (‘foil’ picture, under a *exact* reading of numeric expressions)².

Participants thus had no real choice in primes: they had to pick the only image that corresponded to a possible reading of the sentence. There were two types of primes. In *cumulative primes*, the correct image made the cumulative reading true and the distributive false. Participants would access the cumulative reading in order to give an accurate answer. Conversely, in *distributive primes*, the correct picture only made the distributive reading true, and participants would click on this picture, accessing to the distributive reading of the sentence.

Baseline trials contained sentences without the relevant ambiguity (e.g., “A circle is connected to three squares”). As in primes, they involved a forced choice: the two pictures were such that one (the correct image) corresponded to the unique meaning of the sentence, while the other one made the sentence false.

In target trials, we again used a sentence ambiguous between a cumulative and a distributive reading, but one of the two images corresponded to the cumulative reading and the other one to the distributive reading. Subjects could then follow their preference; they had a choice. Crucially, target trials could be immediately preceded by primes or by baselines. Whenever a target followed a prime, we expected the response to be influenced by the choice that was forced just before: participants should select the distributive image more often after a distributive prime than after a cumulative prime (and vice versa). In contrast, the baseline-target configuration allowed us to measure the rate of cumulative and distributive readings on targets in the absence of immediate priming (i.e. the baseline preference).

² We assume an *exact* reading of numeric expressions. If *at least* readings are also taken into account, most prime trials would allow for a free choice between the two pictures, depending on whether one chooses the *exact* or the *at least* interpretation for the object numeral. Nonetheless, our results show that participants only accessed the exact reading of numerals in this task (see Discussion).

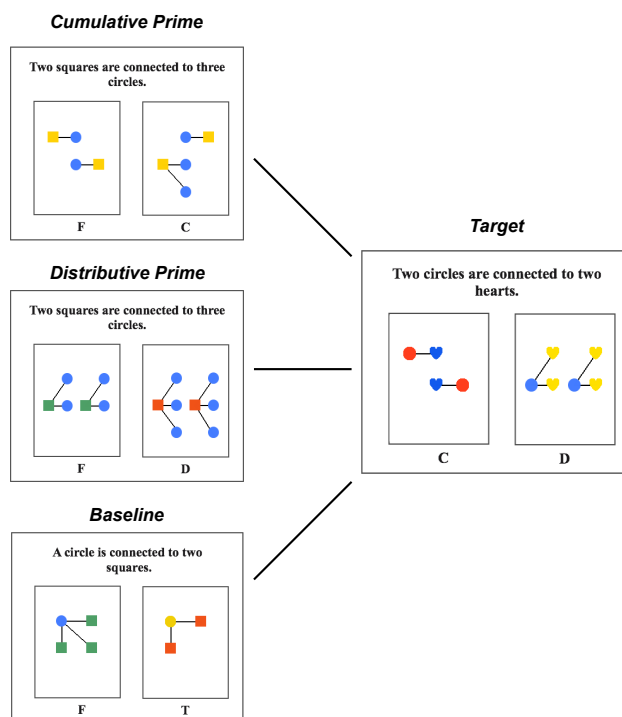


Figure 2.2: Schematic representation of the different types of trials in Experiment 1. Target trials could follow cumulative primes, distributive primes or baselines. Prime and baseline trials were formed by a foil picture (F) and a picture consistent with the sentence (C, D or T, depending on the condition). In target trials the ambiguous sentence was always paired with two consistent pictures (C and D). Image positions (left-right) and numeric expressions in object position ('two'; 'three') were randomized between primes and targets.

2.2.1 Methods

Participants

Sixty participants (29 women, 31 men) were recruited online using Amazon Mechanical Turk (<https://www.mturk.com>). Nine participants were removed from the analysis because they did not report English as their native language and a further two were excluded because their mean response times were lower than 1 second. The remaining 49 participants were taken into account for the analysis.

Materials

Each trial consisted of a sentence and two images appearing together on a computer screen. The sentences were automatically generated using one of two frames (Examples (5) and (6)) and one of two numeric expressions in object position (a. 'two', b. 'three').

- (5) Ambiguous sentences (used in primes and targets):
 - a. Two [shape 1] are connected to two [shape 2]
 - b. Two [shape 1] are connected to three [shape 2]
- (6) Non-ambiguous sentences (used only in baseline trials)³:

³ These sentences may have more than one interpretation due to potential scope ambiguities. Sentence (6) may have an inverse-scope distributive reading equivalent to 'there are two squares such that for each of them a circle is connected to it'. This reading is also true in the situation depicted by the target image (cf. Fig. 2.2), and when accessed it could bias participants towards a "distributive" response. If that were the case, these trials would not be acting as baselines. This potential confound is addressed in Experiment 2.

- a. A [shape 1] is connected to two [shape 2]
- b. A [shape 1] is connected to three [shape 2]

The same predicate ‘be connected to’ was present in all sentences. Shape words were either ‘heart(s)’, ‘square(s)’, ‘circle(s)’, or ‘triangle(s)’, and two different shape words were used in each sentence. Importantly, using different numbers in object position allows for the possibility of an *equal* or *different* combination of numerals in the primes and targets. We generated a list of such sentences with Python, inserting shape words randomly.

Each ambiguous sentence was associated with three images: a foil picture (F) inconsistent with the sentence (i.e. makes all readings of the sentence false), one that made the sentence true under its cumulative reading (C) and another that made the sentence true under its distributive reading (D). For non-ambiguous sentences, there were only two images: a foil picture (F) and a picture that makes the sentence true under its unique interpretation (T; but see fn 3).

Fig. 2.2 presents examples of the different sentences and their corresponding images for each experimental trial (extended version provided in Appendix: Fig. 2.8).

Primes trials were formed by an ambiguous sentence (e.g. (5)) presented with a foil picture (F) and a picture consistent with the sentence under one of its readings (i.e. ‘correct’ picture). While in cumulative primes (Fig. 2.2; upper-left) the correct picture only made the cumulative reading of the sentence true (C picture), in distributive primes (Fig. 2.2; middle-left), the correct image was consistent with the distributive interpretation (D picture). Baseline trials (Fig. 2.2; lower-left) involved a non-ambiguous sentence (e.g. (6)) together with a foil picture (F) and a consistent picture, which corresponds to the unique meaning of the sentence (T picture). Lastly, target trials (Fig. 2.2; right) presented an ambiguous sentence (as primes, e.g. (5)) with two consistent images: one image that matches the cumulative reading (C) and another one that matches the distributive reading (D).

We also include filler trials. They were just like prime trials except that the two pictures involved different visual arrangements (see Fig. 2.3).

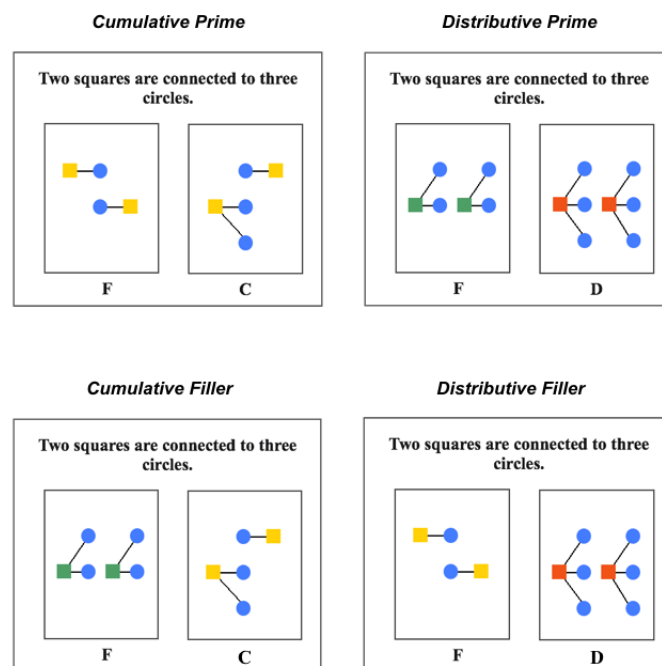


Figure 2.3: Fillers and primes in Experiment 1.

Design

Experimental trials were organized in triplets: each target trial was preceded either by two baselines or by two distributive primes or by two cumulative primes of the same type. Two primes were used instead of one to boost the priming effect and make target trials appear further away from primes associated with a previous target trial.

Each of the trials in a triplet used a different sentence. The two prime/baseline trials always used the same numerals, but with different shapes, word, colors, and different positions on the screen for the correct image. Half of the target trials used the same numerals as their corresponding primes, and the other half used different numerals. In target trials, the position of the image corresponding to the primed reading was either that of the correct image in the first prime or in the second prime.

Target trials therefore appeared in three conditions (cumulative, distributive and baseline), depending on the type of primes in their triplet; four possible number combinations between primes/baselines and targets; and four possible combinations of target image position on the screen.

We crossed these three factors to obtain 48 triplets: 3 (conditions) \times 4 (numeral combinations) \times 4 (image positions). Consequently, the total number of experimental trials was 144. There were a further 48 filler trials randomly inserted between experimental triplets. The total number of trials was 192.

Procedure

Participants were instructed to select the picture that “makes the sentence true”. In cases where both pictures could match the sentence, they were instructed to provide their spontaneous preference. At the beginning of the experiment, they were given an example with a baseline sentence (e.g., “A heart is connected to two squares”).

The experiment was implemented using the Ibx Farm online platform. All participants saw the same set of trials. Experimental triplets and individual fillers were administered in random order to each participant. The presentation paradigm is exemplified in Fig. 2.4.

2.2.2 Results

Data treatment

Accuracy was analyzed for prime and baseline trials. Targets were preceded by two primes or baselines. When participants responded incorrectly to primes, we could not be sure that they had derived the expected interpretation. Thus, targets were discarded as soon as one of the two preceding primes was inaccurate (following the procedure in Raffray & Pickering, 2010). The remaining target responses were coded as 1 when the selected picture corresponded to the distributive interpretation, and as 0 when they matched the cumulative reading. Therefore, the dependent measure in targets was the proportion of distributive choices.

In Experiment 1, the accuracy for baseline trials and both types of primes was higher than 97%. Less than 3% of target responses were removed because of incorrect prime responses. Detailed results for primes, baselines and fillers are provided in the Appendix (Tables 2.1, 2.3 and 2.4).

Analyses procedure

All analyses were carried out in the R programming language and environment (R Core Team, 2014), using the lme4 software package (Bates, Maechler, Bolker, & Walker, 2014). The data and analysis script is available online at <http://semanticsarchive.net/Archive/ThhYTBi0/primingplurals.html>. The responses were analyzed by modeling response-type likelihood

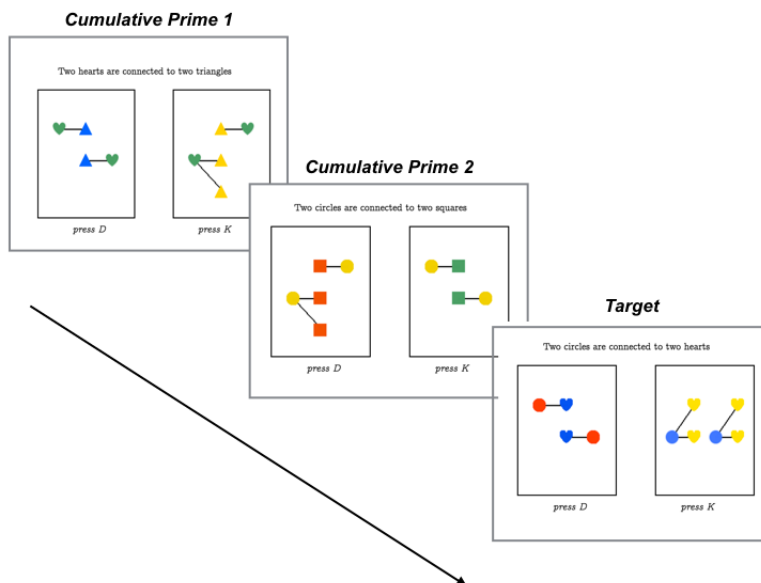


Figure 2.4: Example of the procedure (Cumulative triplet). Immediately after a response to a given trial, the next trial (sentence and pictures) would appear on the screen, whether within or in between triplets. Experimental triplets were thus not identifiable as such.

using logit mixed-effect models (Jaeger, 2008). We included random slopes and intercepts for subjects (within-subject factors, conditions) and random intercepts for items (between-items factors, conditions), following recommendations from Barr, Levy, Scheepers, and Tily (2013). The p -values were obtained by a χ^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 & Bott, 2015; Raffray & Pickering, 2010, among others).

The p -values are reported without correction for multiple comparisons, but the reader can see that Bonferroni corrections will not affect any relevant inference (i.e. when a p -value is lower than the .05 threshold, it is very far from this threshold).

Analyses

Fig. 2.5a shows the mean percentage of distributive responses, i.e. the proportion of target trials in which the distributive image was selected per condition. Figs. 2.6a and 2.6b break this down depending on whether or not the numeral combination was the same in primes and target. Raw means for these conditions are provided in the Appendix (Table 2.1).

We fit a model with a *condition* factor (3 levels: distributive, cumulative, baseline), a *numeric combination* factor (2 levels: same and different than primes/baselines) and their interaction as predictors.

Overall, the proportion of distributive choices in target trials varies significantly across the three conditions (cumulative, distributive and baseline). This is shown by comparing a full model and one which drops the condition factor from the predictors: $\chi^2(1) = 14.9, p < .001$ (Result 1).

The proportion of distributive choices in target trials is significantly higher in the distributive condition than in the cumulative condition. The analysis is the same as above but restricted to the cumulative and distributive conditions: $\chi^2(1) = 11.6, p < .001$ (Result 2).

The proportion of distributive choices in target trials is significantly higher in the distributive

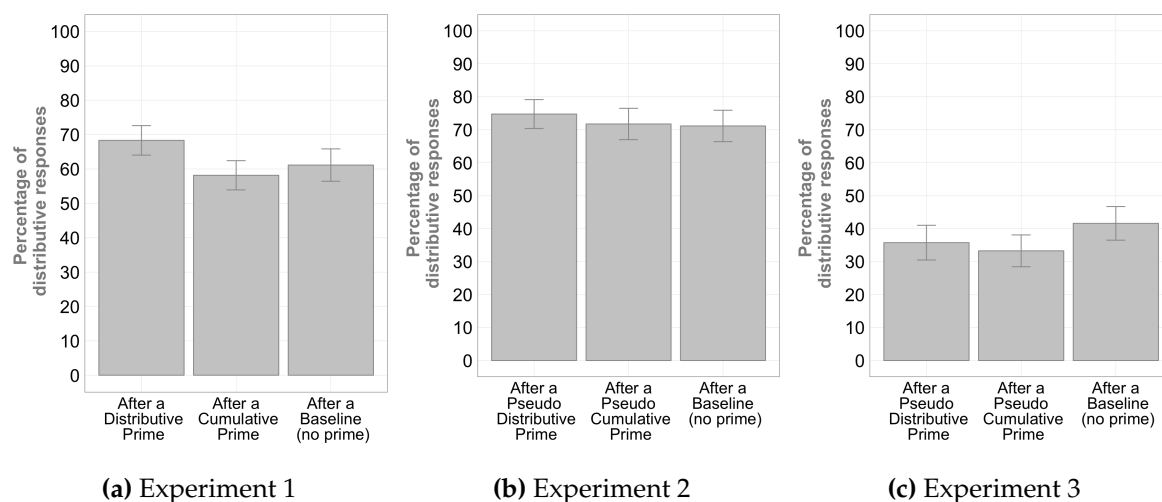


Figure 2.5: Target responses. (a) Mean percentage of distributive responses to targets in Experiment 1 (all numeral combinations); (b) Mean percentage of distributive responses to targets in Experiment 2 (described in forthcoming Section 2.3); (c) Mean percentage of distributive responses to targets in Experiment 3 (described in forthcoming Section 2.3).

condition than in the baseline: $\chi^2(1) = 10, p = .0015$ (Result 3).

The proportion of distributive choices in target trials is not significantly lower in the cumulative condition than in the baseline: $\chi^2(1) = 1.19, p = .27$ (Result 4).

The difference between baseline and cumulative conditions is significantly different from the one between baseline and distributive conditions. This is demonstrated by comparing a model which includes two independent predictors, a coefficient for the cumulative condition and another one for the distributive condition, to a model with a single (numerical) predictor whose value is 1 for cumulative and -1 for distributive conditions (and 0 for baselines). The model comparison shows that it is significantly better to include a *different* coefficient for each potential priming condition ($\chi^2(1) = 8.82, p = .012$), which indicates that the effects of distributive and cumulative primes on target trials are best modelled as having different magnitudes (Result 5).

We did not detect an interaction between Condition and Numeric combinations (same or different): $\chi^2(1) = 2, p = .36$ (Result 6)⁴.

2.2.3 Discussion

General summary

First, the high accuracy rate in prime trials indicates that participants can access both cumulative and distributive interpretations. Result 1 establishes that different primes have different effects on participants' responses to target trials, i.e. that some kind of priming effect is operative. More specifically, the rate of distributive interpretations in target trials is higher when the primes are distributive than when they are cumulative (Result 2). It is also higher with distributive trials than in the baseline condition (Result 3). This suggests that abstract semantic properties such as distributivity and cumulativity might be mentally represented (but see alternatives in subsections below). Given the fact that primes and targets were based on different sentences, no priming effect would be expected unless participants were sensitive to some abstract commonalities within each condition. Note that the sentences used different shapes and that we did

⁴Posthoc analyses in each subgroup of data suggest however that the overall effect of condition is significant when the triplets share the numeric combination ($\chi^2(1) = 10.42, p = .005$), but only marginally so when primes and targets do not share numeric expressions ($\chi^2(1) = 5.2, p = .07$). Given the lack of interaction (Result 6) and the decrease of statistical power, it is difficult to interpret these results.

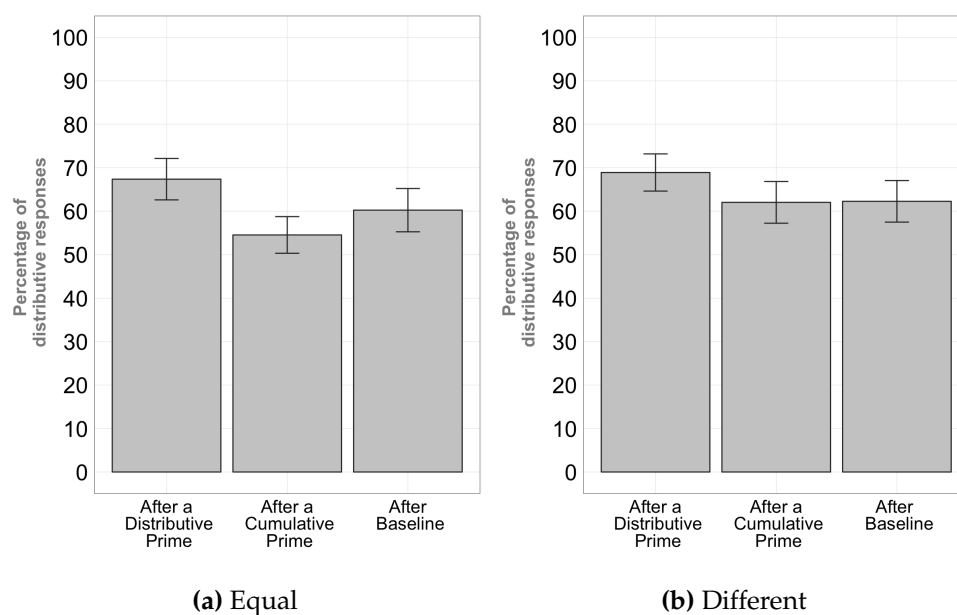


Figure 2.6: Target responses per numeral combination (Experiment 1). (a) Mean percentage of distributive responses when object numeric expressions in primes and targets are the same (‘two’-‘two’ or ‘three’-‘three’). (b) Mean percentage of distributive responses when the numerals in primes and targets are different (‘two’-‘three’ or ‘three’-‘two’).

not find that the presence of a priming effect depended on whether sentences in primes and targets used the same numeric expressions (Result 6).

Result 4 indicates no significant difference between cumulative and baseline conditions: participants’ baseline preference is not biased by exposure to a cumulative prime (i.e. cumulative choices do not increase after cumulative primes). This result suggests the possibility of an *asymmetry* between distributive and cumulative conditions, whereby distributive readings would be more primed by distributive primes (relative to the baseline preference) than cumulative readings are primed by cumulative primes.

Result 5 reports an analysis testing whether the difference between cumulative and distributive conditions could be solely due to the directionality of the effect (i.e. going in opposite directions) or whether these effects are also of different magnitudes. The result finds support for a true asymmetry between the effects.

As mentioned in the introduction, this asymmetry is expected if part of the priming effect is specifically driven by the presence of a distributive operator in the LF of sentences. More specifically, the source of the asymmetry might be the following. Schematically, if we refer to the cumulative LF as α , the distributive LF is of the form $\Delta(\alpha)$. A representation such as α might thus prime not only an α target, but also the α portion of a distributive target. On the other hand, a $\Delta(\alpha)$ prime would be expected to give rise to more priming of a $\Delta(\alpha)$ target than a mere α target. Essentially, the distributivity operator Δ may be the locus of the priming effect, and this would thus be active only with distributive primes. However, alternative interpretations of our results are also available and must be considered.

The possible role of *at least* readings

We will start by taking a closer look at the possibility that participants accessed the *at least*-reading of numeric expressions in the object position (readings paraphrasable by, e.g., *Two circles are connected to at least two squares*), and seeing whether taking this possibility into account would alter the interpretation of our results.

As already noted, in most primes, the availability of *at least* readings would have the effect

of making both pictures correct for one reading or another of the relevant sentence. This is the case for all primes except cumulative primes based on the 2-3 numeral combination, and baseline associated with sentences of the form *A circle is connected to three squares* (modulo changes in shapes and shape words). The very high accuracy on primes (~97%) suggests that participants always picked an *exact*-reading in prime trials (otherwise the F-picture would have sometimes been chosen). This makes us confident that, in general, the readings accessed were *exactly*-readings, so that our interpretation is not undermined by the potential availability of *at least* readings.

A detailed description of the trials –displayed in Fig. 2.8 (Appendix)– will nonetheless allow us to illustrate the possible role of *at least*-readings in the experiment. On the *at least*-construal, both pictures of distributive primes (D and F) are ‘correct’ choices on some reading of the sentence (the D-picture is correct on both the cumulative and the distributive *at least*-reading, and the F-picture is correct on both readings when the 2-2 numeral combination is used, and only on the cumulative reading when the 2-3 numeral combination is used). The fact that participants behave as if they only accessed the *exact*-reading (they always pick the D-picture) might derive from the participants’ tendency to look for a reason to choose one picture over the other. This would pressure them to select an *exact*-reading, since on an *exact* construal, but not an *at least* construal, only one picture can make the sentence true. Thus, our distributive primes might conceivably favour *exact* distributive readings and if they give rise to a priming effect, we need to assess whether what is primed is really the distributive construal (as we assumed) rather than the *exact* reading. Importantly, priming of the *exact* reading *per se* is not sufficient to explain the increased rate of (exact) distributive choices in targets after (exact) distributive primes. This is because, in target trials, both the D-picture and the C-picture actually make an exact reading true: the D-picture makes the sentence true under the *exact* distributive reading, and the C-picture makes it true under the *exact* cumulative reading. The increased rate of distributive choices (selection of the D-picture) cannot be due to the fact that the *exact* reading of numerals might have been primed. It must specifically be due to a priming of distributive readings. We conclude that the interpretation of our main finding is not undermined by the theoretical availability of *at least* readings: we can view the increased rate of distributive responses after distributive primes (compared to cumulative and baseline conditions) as reflecting a priming of distributive readings.

If the possibility of *at least* readings cannot explain the priming of distributivity that we observed, could it nonetheless explain the asymmetry between distributive and cumulative conditions?

On the *at least* cumulative interpretation, *both* pictures in target trials make the sentence true (even though only the C-picture makes the cumulative reading true under an *exact* construal), while on an *at least* distributive interpretation, only the D-picture makes the sentence true. Therefore, while we can interpret an increase of distributive responses after distributive primes –compared to baseline– as reflecting distributive priming, it is less clear that we should expect an increase of cumulative responses (choices of C-picture) after cumulative primes, since even the choice of the D-picture is compatible with a cumulative construal (under the *at least* reading). This might be sufficient to explain the absence of a significant increase in cumulative responses in the cumulative condition, as compared to baseline (Result 4), and, more importantly, why distributive responses after distributive primes are increased to a greater extent than are cumulative responses after cumulative primes (Result 5). This asymmetry, which we interpreted above as suggesting that the priming effect was specifically tied to the presence of a distributivity operator, might then receive an alternative explanation in terms of *at least* readings. We should note, however, that in cumulative primes based on the ‘two’-object phrase, the *at least* cumulative reading makes both pictures good descriptions of the sentence. The fact that participants virtually always picked the C-picture in such cases means that they always accessed the exact interpretation. This, in our view, makes it less likely that they accessed an

at least-reading in the subsequent targets. Overall, given that, in all trials where evidence is available, participants seem to have accessed only exact readings, we conclude that the observed asymmetry between distributive and cumulative primes with respect to baselines is unlikely to be due to the availability of *at least* construals; however, such an alternative explanation cannot be entirely ruled out.

Numeral mismatch in sentence-picture pairs

So far, we have examined two possible interpretations of our data, which explain the effect in terms of a particular type of semantic priming. However, our results could also have an alternative explanation that does not involve priming of semantic representations or properties (such as distributivity or *at least* readings), but of specific verification strategies. Indeed, primes and targets in one particular condition share a property which has nothing to do with whether the reading is cumulative or distributive, but rather with the fact that the object numeral appearing in the sentence *matches* (or not) the number of corresponding shapes in the picture.

Distributive primes force a mismatch between the number expressed in the sentence and the one represented in the picture: a sentence such as “Two squares are connected to three circles” will be systematically paired with a picture containing six circles and not three (see Fig. 2.2). Although participants are indeed forced to access the distributive reading to choose the ‘correct’ picture, they also need to choose between two mismatching sentence-picture pairs. This is not the case for cumulative primes, where there is always a parallel between sentence and picture. Thus, people could be primed to choose (mis)matching pairs in targets because they did so in primes, independently of the underlying semantic representation.

Crucially, such a numeral mismatch effect could also account for the asymmetry between cumulative and distributive primes. Several studies (Patson, 2014; Patson & Warren, 2010, among others) have suggested that numeric expressions are interpreted in a highly specific way, representing the exact quantity of individuals named in the sentence (cf. Introduction). The numeral mismatch in the distributive condition would then require a more complex interpretation of the numeric expression that might not be the one obtained by default. This change (i.e. passing from three to six shapes) could not only require extra cognitive resources but also be specifically primed. The asymmetry between distributive and cumulative effects would be thus the consequence of priming a particular “mismatch” operation.

The results of Experiments 2 and 3, reported below, will appear to be incompatible with such an interpretation.

Baseline rates and inverse-preference effects

One may ask what our results can say about the relative preference for the different readings and how the current findings compare with the current literature in that respect. Indeed, while most baseline rates reported in the literature have shown a dispreference for distributive readings (Brasoveanu & Dotlačil, 2015; Dotlacil, 2010; Frazier, Pacht, & Rayner, 1999; Syrett & Musolino, 2013, among others), the baseline rate in our experiment appears to *favor* distributive interpretations (Fig. 2.5a and Table 2.1). However, our experiment investigated the contrast between distributive and *cumulative* readings, whereas other studies contrast distributive and *collective* interpretations. To illustrate, the sentence “John and Mary wrote two papers” can have a cumulative interpretation, where John and Mary may have written one paper each, a distributive interpretation, where John and Mary wrote two papers each, and also a collective interpretation, where John and Mary wrote the two papers together. While our experiment speaks to the comparison between the first two readings, other studies have been interested in the last two. Moreover, many of these studies differ from ours in the specific methodology, using different tasks (e.g., truth-judgments; Dotlacil, 2010), or different stimuli (e.g., other types of predicates; Brooks & Braine, 1996), and such experimental factors could influence

preferences (Degen, 2013). Crucially, then, experimental designs more similar to ours might offer similar results. This is indeed the case of an early study by Gil (1982). The author compared the availability of distributive and cumulative interpretations in Hebrew and Dutch, using stimuli similar to ours, and found that distributive and cumulative interpretations were equally available.

Hence, our results do not seem to conflict with current knowledge about preferences for distributive readings. Furthermore, we would like to suggest that our findings are actually consistent with the view that there may not be such a thing as a natural derivation rate of distributive readings. We found that the rate of distributive readings could be influenced even within a single experiment from one item to the next, through priming effects, and it is therefore not surprising that the observed rate of distributive reading can vary from one experimental setting to another. One may then ask how fluctuations in baseline rates due to general factors might impact the priming effect: e.g., is the priming effect only present in the specific conditions of the experiments we present, and possibly not in others in which distributive readings would be dispreferred?

As stated in the Introduction, psycholinguistic literature has suggested that baseline rates can indeed influence priming, giving rise to an *inverse-preference pattern* such that structural priming is stronger for dispreferred or infrequent structures (cf. English passive and active sentences; review in Ferreira & Bock, 2006; Pickering & Ferreira, 2008). The inverse-preference pattern leads to the expectation that, in a context where distributive readings are dispreferred, distributive priming should be stronger than cumulative priming. This prediction would hold for experimental setups with low distributive baselines. However, an inverse-preference effect will not arise in experimental settings with a high baseline rate of distributive readings. In the absence of such reverse pattern, the asymmetric priming attested in our experiment has to be seen as an effect that arises *despite* the fact that distributive readings are not dispreferred and not *because* of it. Therefore, the current results may constitute evidence that priming of distributivity could be rather robust to experimental variations.

Visual priming

Finally, an explanation in terms of a *visual* priming effect should be considered. It has been shown that a visual stimulus is more easily processed when it is repeated, changing typical preference patterns and recognition times (non-verbal perceptual priming, Tulving and Schacter 1990, among others). The general similarity between pictures that correspond to the same interpretation might by itself be responsible for a priming effect. This similarity would have to operate at quite an abstract level, independently of the number and type of shapes present in prime and target images (i.e. priming should be driven by more abstract visual commonalities between primes and targets). This interpretation would also need to be quite sophisticated to explain why the priming effect we found is asymmetric. But we cannot rule out that visual priming could contribute to some of our findings. Experiments 2 and 3 were therefore designed to rule out this alternative explanation in a conclusive way.

2.3 Experiments 2 and 3: Control experiments

Experiments 2 and 3 aimed to test the presence of purely visual priming effects in participants' performance. In both cases we modified the design of Experiment 1 by replacing some or all of the sentences in prime trials (but not target trials) without changing the pictures, and we made sure that participants would still be forced to pick the very same picture as in Experiment 1. Crucially, though, the choice of the correct picture no longer involved a distributive reading, but some other kind of reading (see below for details). In other words, in both experiments we eliminated the possibility of any distributive priming (by taking distributivity out of the

picture), but not that of a visual priming effect (by keeping the pictures exactly the same). Examples of the trials are provided in Fig. 2.7.

2.3.1 Methods

Participants

A different group of 40 participants (Experiment 2: 20 women, 20 men; Experiment 3: 20 women, 20 men) was tested for each experiment. None of the participants in Experiments 2 and 3 had done Experiment 1. Participants that were not English native speakers were excluded from the analysis (five participants in Experiment 2 and six participants in Experiment 3). A further two participants were excluded from Experiment 2 because their mean response times were below 1 second.

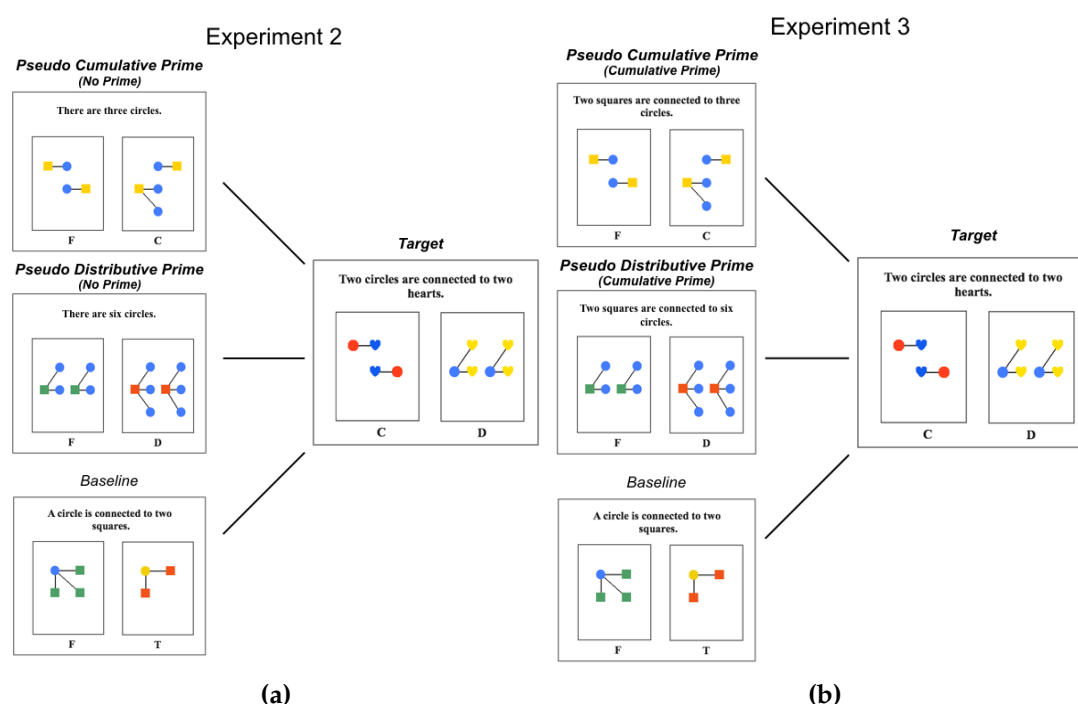


Figure 2.7: Illustration of trials in Experiments 2 and 3. (a) Prime sentences in Experiment 1 (cf. Fig. 2.2) were replaced by unambiguous sentences in Experiment 2. In *pseudo distributive primes*, what was the distributive picture (D) in Experiment 1 is still the only correct picture, but no longer instantiated a distributive reading (\sim *pseudo cumulative primes*). Targets and baselines were the same as in Experiment 1.; (b) Only distributive primes of Experiment 1 where replaced. Sentences in *pseudo distributive primes* were ambiguous between a cumulative and a distributive reading, but were such that the distributive picture (D) of Experiment 1, while still the correct one, now corresponded to the *cumulative* reading of the sentence. *Pseudo cumulative primes* were similar to cumulative primes in Experiment 1 (same for targets and baselines).

Materials

In Experiment 2, we replaced ambiguous sentences in primes (i.e. former cumulative and distributive primes) with unambiguous sentences, keeping the same images. In each resulting trial, the sentence was true with respect to one image and not the other and, importantly, we made sure that the 'correct' image was always the same as in the corresponding trial of Experiment 1. In other words, participants were led to choose the same picture as in Experiment 1, but for a different reason (see Fig. 2.7a and example sentences in (7)).

- (7) Experiment 2
- a. Pseudo cumulative prime: There are two/three [shape2]
 - b. Pseudo distributive prime: There are four/six [shape2]

We will call the resulting trials *pseudo distributive* and *pseudo cumulative* primes, depending on the ‘correct’ picture in the trial.

In Experiment 2, we do not expect any priming effect, since the sentences are unambiguous and the relevant trials do not instantiate cumulative or distributive readings. What we now call a pseudo distributive prime (respectively pseudo cumulative prime) is a trial (in a triplet) obtained from a distributive (respectively cumulative) prime from Experiment 1 in the way just described. Baselines and targets were identical to the ones in Experiment 1 (see Fig. 2.7a). The use of the same non-ambiguous sentences across experiments served to validate the baselines in Experiment 1 (cf. Footnote 3): in Experiment 2, responses in all target trials can now all serve as baselines, whatever the condition is.

In Experiment 3, only the sentences of the distributive primes of Experiment 1 were changed, giving rise to *pseudo distributive primes*. The new sentences were also ambiguous between a cumulative and a distributive reading, but now only the *cumulative* reading corresponded to one of the two images, and furthermore, this ‘correct’ image was the same as in the corresponding *distributive* prime in Experiment 1. In other words, the pictures that instantiated the distributive reading of the associated sentence in Experiment 1 primes now instantiates the cumulative reading of another sentence, generating only cumulative primes (see Fig. 2.7b and example sentences in (8)).

- (8) Experiment 3
- a. Pseudo cumulative prime: Two [shape1] are connected to two/three [shape2]
 - b. Pseudo distributive prime: Two [shape1] are connected to four/six [shape2]

Although Experiment 3 has *only* cumulative primes, for the ease of comprehension we will distinguish ‘pseudo distributive’ primes from ‘cumulative’ primes: a pseudo distributive prime in Experiment 3 is a trial that was obtained from a distributive prime of Experiment 1 in the way just described, and a pseudo cumulative prime of Experiment 3 is a trial that is identical to a cumulative prime of Experiment 1.

Design and procedure

The design for Experiments 2 and 3 was similar to that of Experiment 1. There were the same number of conditions (cumulative, distributive and baseline) and therefore items. The stimuli were obtained by minimally modifying the material from Experiment 1 as described above.

As in Experiment 1, a single list of trials was used for each experiment, following the procedure exemplified in Fig. 2.4. This list was then administered to each participant in a random order using Ibx Farm.

2.3.2 Results

Responses were analyzed as in Experiment 1. The accuracy in (pseudo) primes was above 93% for both Experiments 2 (M=94.9) and 3 (M=93.4). Complete data for all the trials can be found in Tables 2.3 and 2.4 (Appendix). As in Experiment 1, the dependent variable was the response given in target trials (choice between the distributive and cumulative pictures). Figures 2.7a and 2.7b show the rate of distributive responses in targets for each condition in both Experiment 2 and Experiment 3.

The overall analysis revealed no significant difference between conditions (3 levels: distributive, cumulative and baseline) in Experiment 2 ($\chi^2(1) = 1.21, p = .54$). However, a main effect of Condition was found in Experiment 3 ($\chi^2(1) = 9.43, p = .0089$). Further analyses

revealed that this effect is driven by the baseline, which has a significantly higher proportion of distributive choices than both cumulative and distributive conditions ($\chi^2(1) = 7.80, p = .005$, and $\chi^2(1) = 5.65, p = .017$). No statistical difference in the proportion of distributive choices was found between cumulative and distributive conditions ($\chi^2(1) < 1, p = .69$).

These results suggest that the priming effect found in Experiment 1 is not driven solely by visual priming, or else the results would not have been altered in these new control experiments (especially Experiment 2). A more direct test of the hypothesis is obtained by directly comparing the control experiments to the original experiment. A significant interaction is obtained between Experiment and Condition, both when comparing Experiments 1 and 2 ($\chi^2(1) = 6.66, p = .035$) and when comparing Experiments 1 and 3 ($\chi^2(1) = 19.8, p < .001$).

2.3.3 Discussion

As expected, the pattern of results found in Experiment 1 was not reproduced in Experiments 2 and 3. Crucially, the difference between cumulative and distributive conditions was significantly reduced. The specific effect of Condition found in Experiment 1 was altered in both Experiments 2 and 3, which shows that the priming in Experiment 1 was not a purely visual effect –the effect disappears when priming sentences are changed, while keeping everything else constant, including the ‘correct’ pictures in the (pseudo) primes.

It is important to note that in Experiment 2, all conditions give rise to a similar quantity of distributive responses, showing that our baseline items in Experiment 1 (which were not changed in Experiment 2) were appropriate. Such results not only alleviate the potential problem noted in Footnote 3 (i.e. potential ambiguity of baseline sentences is not at play), but also support the idea that asymmetric results in Experiment 1 cannot be explained in terms of an inverse-preference effect. An inverse-preference pattern would explain the results of Experiment 1 if distributive readings were dispreferred in *complete* absence of priming. Since in Experiment 2 participants prefer distributive readings across the board (as in baselines of Experiment 1), an interpretation in these terms proves not to be on the right track.

Strikingly, in Experiment 3, the overall rate of distributive responses is much lower than in Experiment 2. In fact, *posthoc* analyses reveal a main effect of Experiment between Exp 2 and Exp 3 ($\chi^2(1) = 23.0, p < .001$): the repeated use of trials that force a cumulative reading in Experiment 3 results in a general increase of cumulative responses in targets, compared with Experiment 2. On top of this *global* bias –which can be seen as adaptation, learning or even long-lasting priming (Fine et al., 2013; Kuperberg & Jaeger, 2016; Pickering & Ferreira, 2008)–, Experiment 3 also displays a *trial-to-trial* cumulative priming. Indeed, the rate of distributive responses was lower in the cumulative and (fake) distributive conditions than in baselines, which means that primes that force a cumulative reading (which in Experiment 3 are all the primes) yield an increase of cumulative choices in targets.

These results allow us to rule out the alternative interpretation based on a numeral mismatching effect. Unlike Experiment 1, Experiments 2 and 3 always present “matching” (pseudo) primes: the number of shapes in the pictures matches the number appearing in the sentence in all prime trials (cf. Fig. 2.7). If the match/mismatch interpretation is correct, we would then expect a strong preference for matching responses across the board in these experiments, that is a preference for cumulative responses. However, the rate of cumulative responses does not increase from Exp 1 to Exp 2 (in fact it non-significantly decreases: $\chi^2(1) = 3.35, p = .067$). It does increase in Exp 3, but this is arguably because then all primes are genuine cumulative primes (not only matching primes).

The fact that cumulative readings influence preference patterns in targets (cf. Experiment 3) appears to conflict with our suggestion that the priming effect detected in Experiment 1 was asymmetric, i.e. that distributive priming is stronger than cumulative priming, if there were cumulative priming at all. There is, however, no real contradiction. If we assume that distributive priming is, all else being equal, stronger than cumulative priming (specifically in

the case of *trial-to-trial* priming effects, i.e. effects of a trial on the next trial), but that cumulative priming nevertheless exists, we expect an overall increase of cumulative responses when we replace all (pseudo) primes with cumulative primes (Experiment 3 vs. Experiment 2). In Experiment 3, the fact that we observe a significant difference between targets after primes (pseudo distributive and pseudo cumulative) and after baselines is not unexpected given that there were in total many more cumulative trials in Experiment 3 than in Experiment 1, and no genuine distributive primes getting in the way.

Our results are therefore consistent with the view that a) there is an asymmetry of strength in *trial-to-trial* priming between cumulative and distributive primes (i.e. distributive primes have a stronger influence than cumulative primes on the targets that immediately follow them), and b) the presence of cumulative primes throughout an experiment affects the *global* rate of cumulative responses. Overall, the results of Experiments 2 and 3 are therefore not inconsistent with the conclusion from Experiment 1 that distributive priming is stronger than local cumulative priming. Note that Experiment 1 is the only one with a direct bearing on the existence of such an asymmetry.

2.4 General Discussion

Experiment 1 provided evidence that the cumulative vs. distributive ambiguity can give rise to a priming effect. Experiments 2 and 3 strengthened this interpretation by ruling out the possibility that our results in Experiment 1 were entirely driven by visual priming. Furthermore, these control experiments also allowed us to exclude an alternative interpretation in terms of a numeral matching priming, since such effect would not explain the difference in target responses between Experiments 2 and 3.

In Experiment 1, we also detected an asymmetry between cumulative and distributive primes: while distributive primes give rise to an increase of distributive responses in targets compared to baseline trials, this does not occur (at least not to the same extent) with cumulative primes. Although this might resemble the inverse-preference effects documented in the literature, our results do not allow us to conclude that specific preference patterns are at the origin of the asymmetry (i.e. distributive interpretations are actually *preferred* in baseline targets). Instead, this asymmetry suggests that distributive readings involve a different derivation from cumulative interpretations. This difference might be well explained by the presence of a distributivity operator in the relevant cases, which would be responsible for the priming effect we uncovered. However, distributive interpretations may also require a particular interpretation of numeric expressions or specific attentional resources, not shared by cumulative readings. Any of these might well be at the origin of the asymmetric effect.

Nonetheless, the (partly *posthoc*) comparison of our two control experiments suggests that the repeated exposure to cumulative primes in Experiment 3 –together with an absence of distributive priming– led to a cumulative priming effect. These two findings (the asymmetry between distributive and cumulative priming in Experiment 1, and the evidence that cumulative priming took place in Experiment 3) are compatible with the view that while priming effects specifically linked to distributivity might be stronger, both readings can give rise to detectable priming effects.

To conclude, the fact that the cumulative vs. distributive ambiguity can give rise to priming effects across different sentences suggests that this distinction is at play during parsing. One possible interpretation for the asymmetry detected in Experiment 1 is that part of the priming effect we detected is specifically due to the distributivity operator or whatever mechanism is responsible for distributive readings.

2.5 Appendix

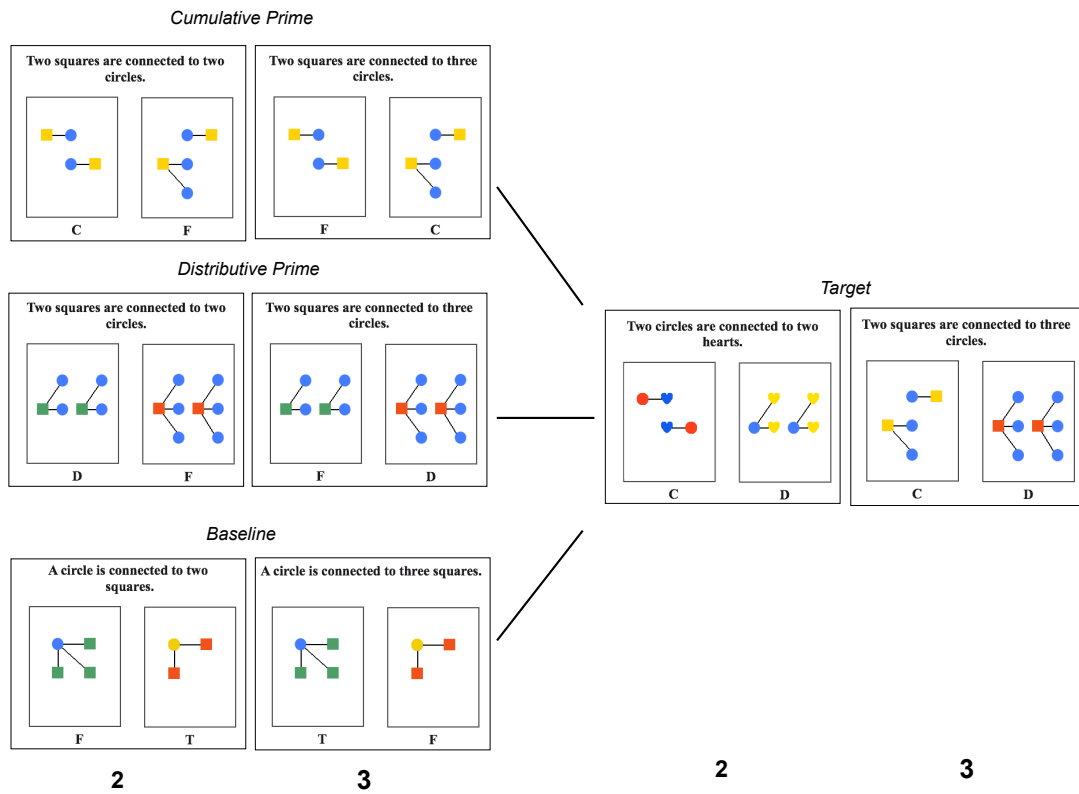


Figure 2.8: Detailed description of trials in Experiment 1

Table 2.1: Raw means for target trials in Experiment 1, 2 and 3.

	Condition	Mean Distributive Choices	SE
Experiment 1	Baseline	61.15	4.71
	Cumulative	58.17	4.25
	Distributive	68.33	4.29
Experiment 2	Baseline	71.15	4.76
	Cumulative	71.74	4.75
	Distributive	74.75	4.38
Experiment 3	Baseline	41.59	5.10
	Cumulative	33.27	4.83
	Distributive	35.75	5.27

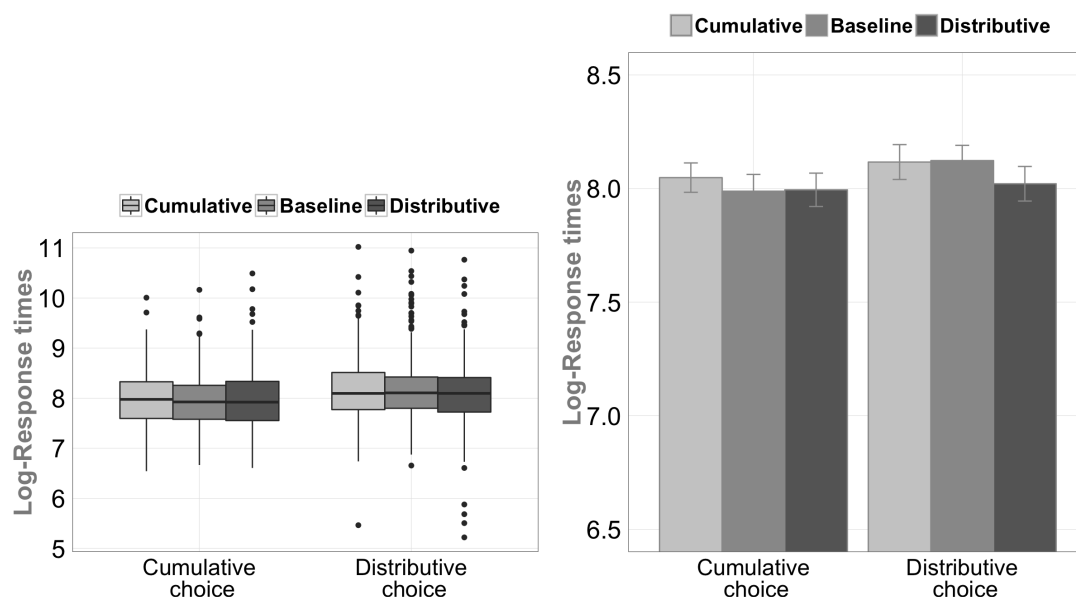


Figure 2.9: Response Times (log-transformed) in target trials (Experiment 1). The color code indicates the condition (cumulative, baseline and distributive). Data is factored according to the type of response in targets, distinguishing between distributive and cumulative choices.

Table 2.2: Response times for target trials in Experiment 1, 2 and 3.

	Condition	Mean RTs (ms.)	SE
Experiment 1	Cumulative	3922.5	196.8
	Baseline	3897.8	211.7
	Distributive	3882.6	187.8
Experiment 2	Cumulative	3648.0	252.0
	Baseline	3595.7	226.8
	Distributive	3557.6	386.5
Experiment 3	Cumulative	3469.9	330.6
	Baseline	3775.3	459.4
	Distributive	2970.7	211.5

Table 2.3: Raw means for prime trials in Experiment 1, 2 and 3.

	Condition	Mean Accuracy (%)	SE
Experiment 1	Baseline	97.51	0.62
	Cumulative	97.32	0.57
	Distributive	97.19	0.53
Experiment 2	Baseline	94.35	1.91
	Cumulative	95.67	1.16
	Distributive	94.86	1.35
Experiment 3	Baseline	94.03	1.39
	Cumulative	93.47	1.59
	Distributive	92.65	1.80

Table 2.4: Raw means for filler trials in Experiment 1, 2 and 3.

	Condition	Mean Accuracy (%)	SE
Experiment 1	Cumulative	96.01	0.90
	Distributive	95.66	0.82
Experiment 2	Cumulative	89.20	2.49
	Distributive	90.73	1.97
Experiment 3	Cumulative	92.58	1.92
	Distributive	91.91	2.42

3 | Priming the collective/distributive contrast for gradable adjectives*

Maldonado, M., Chemla, E. & Spector, B. Revealing abstract semantic mechanisms through priming: the distributive/collective contrast. Submitted.

Abstract

Sentences such as *The bags are light* allow both collective (i.e. they are light *together*) and distributive interpretations (i.e. *each* bag is light). We report the results of two experiments showing that this collective/distributive contrast gives rise to priming effects. These findings suggest that the abstract interpretative mechanisms posited to account for this ambiguity are represented as such during comprehension, independently of the specific verification strategies associated with each reading.

3.1 Introduction

In the last thirty years, priming has served to identify the abstract representations that people construct when producing or comprehending language (Branigan & Pickering, 2017; Pickering & Ferreira, 2008, for reviews). This type of priming is known as *structural priming* and it occurs when the processing of a structure is facilitated after the same structure has been recently processed. While structural priming has often been associated with *syntactic priming*, recent studies have revealed that priming methods also serve to tap into abstract semantic mechanisms at play during the interpretative process (L. Bott & Chemla, 2016; Feiman & Snedeker, 2016; Maldonado, Spector, & Chemla, 2017; Raffray & Pickering, 2010).

Semantic theories have proposed the existence of invisible operations to derive specific sentence interpretations. For example, a silent distributivity operator (*D* operator) has been proposed to explain why sentences such as “Two boys hold three bags” can have not only a basic *cumulative* reading (e.g., Two boys hold three bags *in total*) but also a distributive interpretation (e.g., Two boys hold three bags *each*). Its meaning roughly corresponds to that of *each* in English (Champollion, to appear; Link, 1983). When modified by the *D* operator, the VP ‘hold three bags’ applies to each atomic member of the plural subject, so each boy is allowed to hold three bags (i.e. the bags can *covary* with each boy). Distributive readings are thus explained by postulating the presence of this *D* operator in the semantic representation.

Using a priming paradigm, Maldonado, Chemla, and Spector (2017) have recently shown that this cumulative/distributive contrast gives rise to priming effects. Specifically, they found evidence for an asymmetric distributive priming, suggesting that an abstract mechanism such as the one proposed by semanticists is indeed at play during the comprehension of these ambiguities and can be primed.

We wish to thank Amir Anvari, Milica Denic, Manuel Kriz, Jeremy Kuhn, Salvador Mascharenas, as well as the audiences at Pallmyr (UCL), Linguae (Institut Jean-Nicod) and Linguistic Evidence 2018 (Tübingen). The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.313610 and from the Agence Nationale de la Recherche (Grants ANR-10-LABX-0087 IEC, ANR-10-IDEX-0001-02 PSL and ANR-14-CE30-0010-01 TriLogMean)

Importantly, the optional insertion of the *D* operator has been proposed to account not only for the cumulative/distributive contrast but also for every sentence that can optionally have a distributive interpretation. In this paper, our goal is to extend these results to what is thought to be another instantiation of the same operator: the collective/distributive ambiguity illustrated in (1) and (2):

- (1) The bags are heavy.
- a. COLLECTIVE READING
The bags together are heavy, without each bag necessarily being heavy.
 - b. DISTRIBUTIVE READING
Each bag is heavy (and the bags are heavy in total as well).
- (2) The bags are light.
- a. COLLECTIVE READING
The bags together are light (and each bag is light as well).
 - b. DISTRIBUTIVE READING
Each bag is light, without the bags necessarily being light in total.

In their collective reading, (1) and (2) are true as long as the predicate can denote a property of the plural subject as a whole, without necessarily being true of each individual member. Distributive readings, instead, entail that the predicate is true of each atomic member of the plural subject. VPs that present this ambiguity, such as ‘heavy’ or ‘light’, are called ‘mixed’ predicates (Link, 1983; Scha, 1984; Schwarzschild, 2011)¹.

Note that, despite being distinguishable, collective and distributive readings of (1) and (2) are not logically independent: one reading entails the other. A scenario that makes the distributive reading of (1) true (i.e. each bag is heavy) also makes the collective reading true. The distributive interpretation *entails* the collective interpretation. This entailment is asymmetric: the collective reading of (1) can be true while the distributive reading is false. Changing the polarity of the adjective switches the direction of the entailment (see Table 3.1). In other words, the entailment is the opposite one for two antonyms (e.g., ‘heavy’ and ‘light’, see Table 3.1). Two gradable adjectives are antonyms (e.g., ‘heavy’ and ‘light’) when they use the same dimension (scale), but they have an opposite ordering of degrees on this dimension (Kennedy, 1999; Rett, 2008, a.m.o). In principle, the difference between positive and negative adjectives can be thought of as the difference between an increasing or a decreasing ordering of degrees. This entailment pattern arises for most *adjectival* mixed predicates (e.g., expensive/cheap; noisy/quiet).

Collective interpretations of (1) and (2) seem to be the result of just applying the plural subject to the predicate, whereas distributive readings are thought to arise by inserting the covert *D* operator. That is, the ambiguity between collective and distributive interpretations of adjectival predicates is explained analogously to the distributive/cumulative contrast tested by Maldonado, Chemla, and Spector. If the same operator is required to derive optional distributive readings across different sentences and predicates, we would expect to see priming effects at play for sentences such as (1)/(2) as well as in the cases discussed in the previous priming study.

Finding priming effects related to the collective/distributive ambiguity would provide further evidence for the existence of an abstract mechanism to derive distributive readings. Moreover, the use of adjectival predicates brings two important advantages. First, as observed, distributive and collective interpretations can be weak or strong depending on the polarity of

¹There is a question in plural semantics of whether or not all gradable adjectives give rise to the collective/distributive ambiguity. Some gradable adjectives have been traditionally considered to be “stubbornly-distributive” in that they do not seem to admit collective interpretations (Schwarzschild, 2011). Recent evidence, however, has challenge this hypothesis (Scontras & Goodman, 2017). Given that the predicates used in these experiments are undoubtedly ambiguous, we will not address this discussion here.

the adjective (cf. Table 3.1). Consequently, ‘mixed’ adjectival predicates allow us to test, for the first time, priming of specific readings independently of logical strength: *weak* distributive readings might prime *strong* distributive reading, while before strong distributive could only be related to strong distributive readings. The entailment pattern of gradable adjectives implies per se that logical form and strength can be distinguishable. However, here we can test whether these two aspects of meaning are also processed separately during parsing.

Furthermore, sentences such as (1) and (2) allow us to dissociate the processing of distributive readings from verification strategies that are not inherent to distributivity. In psycholinguistics, the collective/distributive ambiguity has been mostly investigated by testing transitive sentences such as “The boys are painting a castle” (Brasoveanu & Dotlačil, 2015; Frazier, 1999; Syrett & Musolino, 2013, a.m.o.). Distributive interpretations were isolated here by presenting participants with scenarios where the object co-varies with each member of the plural subject. In the example above then, the distributive scenario would involve a different castle per boy. Participants may use a verification strategy specific to distributive interpretations, based on checking for covariation of the object with respect to the subject. The processing pattern attributed to distributive interpretations therefore confounds verification strategy with mere semantic interpretation. Mixed adjectival predicates allow us to isolate distributivity from covariation (i.e. and more generally, to the strategy required to verify whether a certain reading is true or false), and therefore to remove this confound.

	Collective reading		Distributive reading
POSITIVE ADJECTIVE: <i>The bags are heavy</i>	The bags are heavy together	⇐	Each bag is heavy
NEGATIVE ADJECTIVE: <i>The bags are light</i>	The bags are light together	⇒	Each bag is light

Table 3.1: Entailment relation between readings. Distributive readings asymmetrically entail collective readings of positive adjectives (e.g., ‘heavy’). The reverse pattern is attested for negative adjectives (i.e. ‘light’). The distributive interpretation is the *strong* reading for sentences involving positive adjectives, and the *weak* interpretation for sentences involving negative adjectives (and the other way around for collective readings).

3.2 Experiment 1

3.2.1 Methods and materials

We used a sentence-picture matching task where participants had to read a sentence and match it with one of two pictures (Maldonado, Chemla, & Spector, 2017; Raffray & Pickering, 2010, a.m.o)². In experimental trials, the sentence involved adjectival predicates 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 one reading of the sentence true (whether this reading is the collective or the distributive one depends on the polarity of the adjective, Table 3.1), 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 Figure 3.1).

Primes were designed to force one specific sentence interpretation. There were two types of primes: *Collective primes* displayed a foil and a weak *collective* picture, so participants would click

²Materials, data and analyses for both experiments are provided in https://osf.io/7acwu/?view_only=fc192f0cf4524209babf3a5cf8f48f0e

on the collective picture and access the collective reading of the sentence. *Distributive primes* displayed a foil and a weak *distributive* picture, forcing participants to access the distributive reading. **Targets** could also be either *collective* or *distributive*. They displayed a weak picture, compatible with the collective or the distributive reading depending on the condition, and a ‘blur’ picture. Participants were instructed to select the ‘blur’ option if they felt that the overt picture was not a sufficient match for the sentence (a method modeled from the “covered picture task”, Huang et al., 2013). Table 3.2 illustrates how target responses are indicative of a choice between a collective and a distributive interpretation.

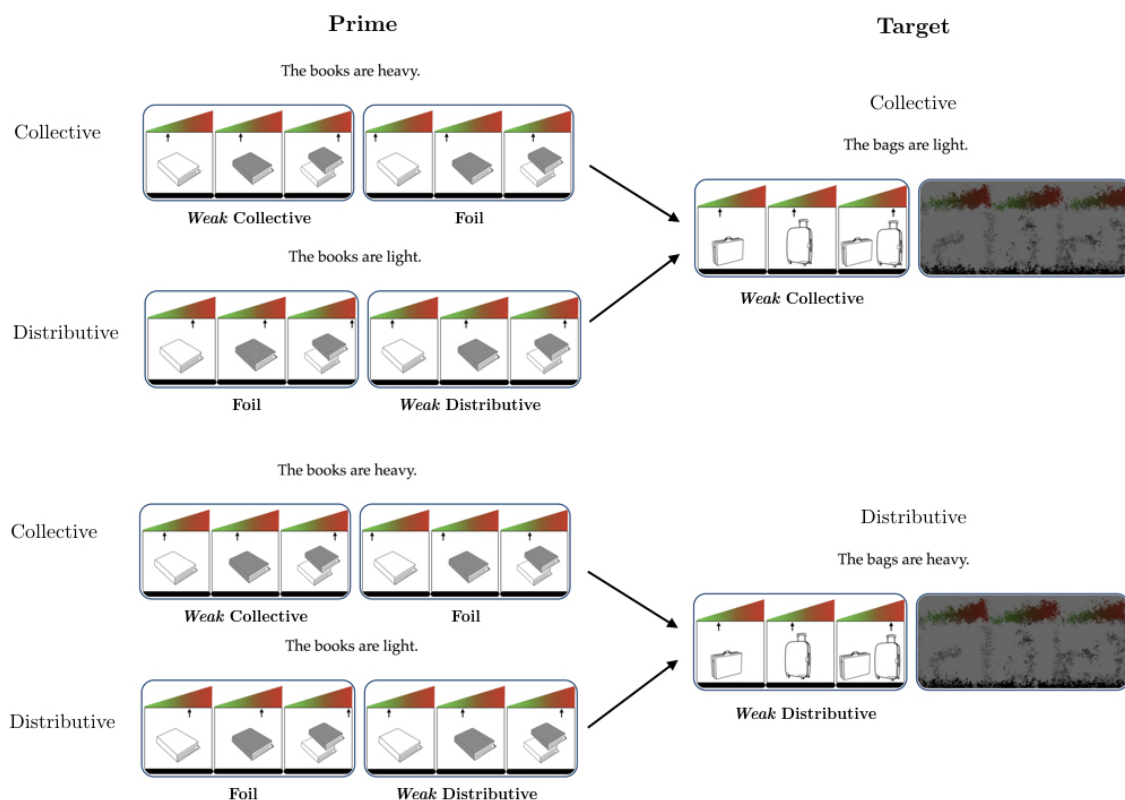


Figure 3.1: Illustration of experimental trials. Each experimental trial involved an ambiguous sentence presented with two pictures. In both primes and targets, sentences were constructed using the frame *The [plural exemplar] are [predicate]*. Predicates were selected among three possible predicate pairs, depending on the scale dimension: light/heavy, cheap/expensive and noisy/quiet. Exemplars varied depending on the predicate pair. Collective primes and targets always displayed sentences involving positive adjectives; distributive primes and targets involved negative adjectives. Prime trials combined the ambiguous sentence with two overt pictures; whereas targets presented an overt *weak* picture and a ‘blur’ picture. Pictures displayed three scales, which could measure degrees of various dimensions (weight, price or sound intensity). Values at the green portion of the scale represented low degrees; values at the red portion represented high degrees.

Targets immediately followed prime trials. After being biased towards one specific sentence interpretation in primes, participants were expected to select more often a picture compatible with this same interpretation in targets, independently of the target condition. For example, collective primes should lead to a greater proportion of overt responses in collective targets and of ‘blur’ responses in distributive targets. Priming of semantic interpretation would then be observed as a main effect of Prime condition in target responses. Given that the picture compatible with the primed reading is not the same across target conditions, we control for visual priming.

The four possible prime-target combinations were present in the experiment. There were two primes of the same condition preceding each target (cf. Maldonado, Chemla, & Spector,

	COLLECTIVE TARGET (Positive adjective) <i>e.g., The bags are heavy</i>	DISTRIBUTIVE TARGET (Negative adjective) <i>e.g., The bags are light</i>
Collective reading	Overt weak, collective picture	'Blur' picture
Distributive reading	'Blur' picture	Overt weak, distributive picture

Table 3.2: Pictures compatible with each reading in target conditions. Image selection in targets is an indicator of the reading participants have accessed. In collective targets, the collective reading is made true by the overt *weak* picture, whereas the distributive reading is only compatible with the 'blur' option. Distributive targets display the reverse pattern: the distributive reading is made true by the overt option, and the collective reading is only compatible with the 'blur' scenario. Since participants were instructed to select the 'blur' option only when the overt (weak) picture was not satisfying (as a last resort), the selection of the 'blur' option was taken to be an indicator of strong interpretations (participants were unsatisfied with a choice that would be compatible with the weak interpretation).

2017), forming experimental *triplets*. Primes and target within one experimental triplet could use predicates from the same or different degree dimension (i.e. height, price, volume), resulting in matching or mismatching predicate conditions. The left/right position of the 'weak' image was counterbalanced (the 'weak' image could be in the same location in primes and target or in a different one).

As a result, the experimental design consisted of four fully-crossed factors to obtain 48 experimental items: 2 (PRIME CONDITION) \times 2 (TARGET CONDITION) \times 2 PREDICATE CONDITION \times 2 (WEAK IMAGE POSITION) \times 3 items (two primes, one target).

A further 64 controls trials were randomly inserted between triplets. These controls were designed to highlight both 'blur' and overt pictures as possible correct responses, preventing participants from developing verification strategies based on the type of picture (see Supplementary Materials for full description).

We recruited 54 participants using Amazon Mechanical Turk, all of whom reported English as their native language. After application of a pre-determined exclusion criterion of 75% accuracy on control trials, 33 participants were considered for the analyses.

3.2.2 Results and discussion

Analyses were carried out in the R programming language and environment (R Core Team, 2014), using the lme4 software package (Bates et al., 2014). The responses were analyzed by modeling response-type likelihood using logit mixed-effect models (Jaeger, 2008), keeping the random structure maximal when possible; p -values were obtained by a χ^2 likelihood ratio test comparing each model with a simpler one in which the relevant predictor was removed (Barr et al., 2013). The dependent measure was the log-odds of choosing a distributive over a collective response on target trials, after each type of prime (see Table 3.2 for reading-picture correspondences). Target responses that were not preceded by two correct prime responses ($\sim 10\%$) were excluded from the analyses.

Figure 3.2a illustrates the mean percentage of distributive responses after accurate primes. Figure 3.2b separates the results depending on PREDICATE CONDITION (matching or mismatching dimension between primes and targets). A first analysis reveals a significant effect of Prime Condition ($\chi^2 = 30.76$; $p < .001$), such that the rate of distributive responses was overall higher after distributive primes than after collective primes. The effect was still significant when the analysis was restricted to mismatching cases ($\chi^2 = 16.028$; $p < .001$), indicating that this priming effect is not tied to the specific predicates. We also note that these effects could not be explained by visual priming: as visible in Figure 3.1, the correct picture in the different prime

conditions are the same (only the sentence changes).

These findings suggest that priming of interpretation is at play: participants' choices in targets appear to be influenced by the reading that was forced in prime trials. So far, however, one cannot tell whether the priming was triggered by all types of primes, or by solely, say, distributive primes. Maldonado, Chemla, and Spector, for instance, found an asymmetry between distributive and cumulative priming, such that distributive primes influenced target selection, whereas cumulative primes behaved just like baselines (i.e. targets after no prime). Such inquiries require a baseline rate of responses in Targets in the absence of primes, absent here.

Finally, we should note an alternative explanation of the results in terms of a 'verification strategy priming'. To illustrate (cf. Figure 3.1), in distributive primes, participants may note that only the values of the first two first scales (those concerning single objects) matter. If so, participants may be biased to decide on the basis of only these two scales in subsequent targets (i.e. 'check-two-scales' strategy), effectively leading to an increase of distributive responses. Similarly, collective primes may lead participants to focus on the third scales (concerning all objects together), which would lead to accept the collective image without even noticing that it makes the distributive reading false (i.e. 'check-one-scale' strategy). In short, readings may trigger specific verification strategies, which may lead to shallow acceptance in Targets, mimicking the effect of priming of readings we are interested in. Experiment 2 was designed to rule out this alternative explanation and to determine from which of the prime(s) condition(s) the effect follows.

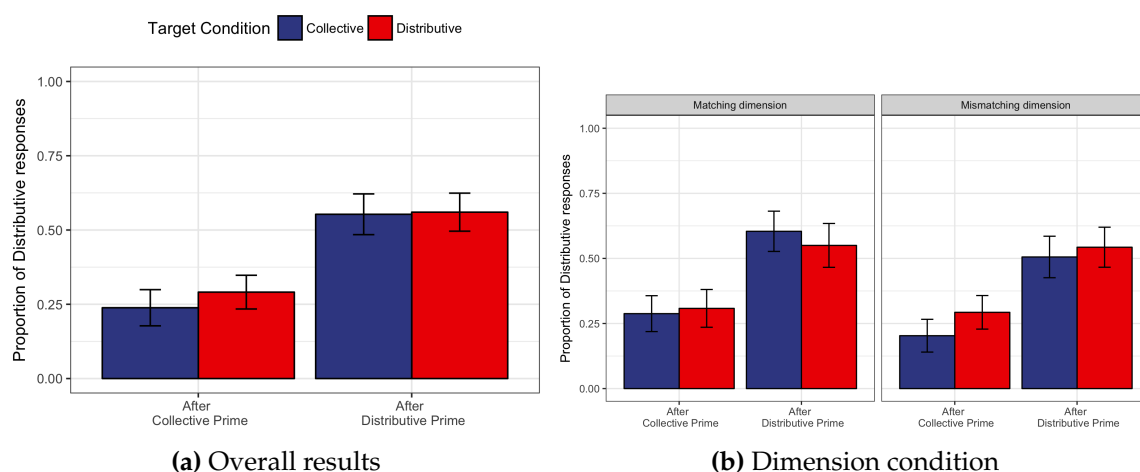


Figure 3.2: Target results in Experiment 1. (a) Mean proportion of distributive responses in collective and distributive targets after each prime condition. Distributive responses correspond to the selection of the overt picture in distributive targets (red) and of the 'blur' picture in collective targets (blue). (b) Mean proportion of distributive responses per predicate dimension. In the matching dimension, primes and targets instantiate predicates of the same dimension (e.g., heavy/light); in the mismatching dimension, predicates in primes and target belong to different dimensions (e.g., expensive/light).

3.3 Experiment 2: Baseline inclusion

3.3.1 Methods and materials

Experiment 2 was similar to Experiment 1 except that baseline experimental triplets were added, in which two *pseudo-primes* replaced the two primes, see an example in Figure 3.3.

Pseudo-primes were associated with one picture that made the sentence true (correct picture), and one that made it false (foil picture), and they involved sentences that do not instantiate

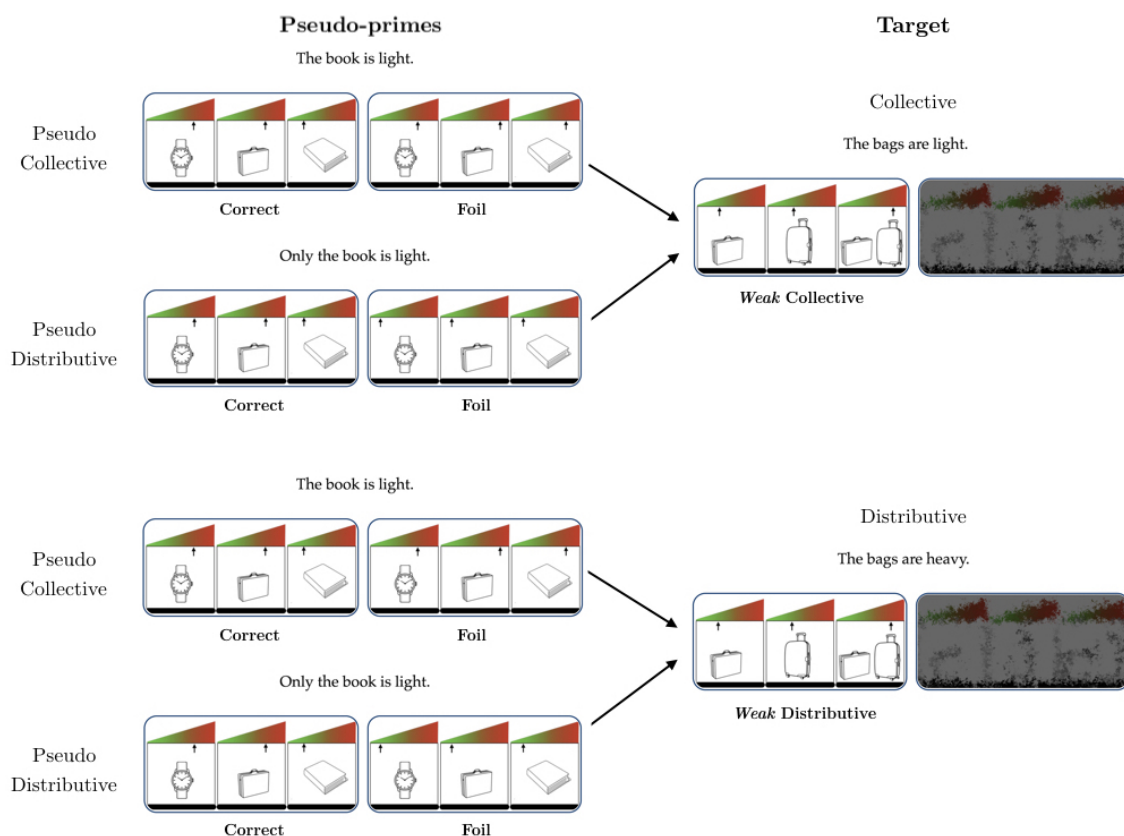


Figure 3.3: Pseudo-prime–Target combinations in Experiment 2. On top of the four prime–target combinations of Experiment 1 (see Figure 3.1), Experiment 2 included four *pseudo-prime*–target combinations. All pseudo-primes involved unambiguous sentences and forced the selection of a picture with ‘mismatching’ scales. Participants had to choose the same type of image as in primes, but, since the sentence was not ambiguous, they were not biased towards one specific interpretation. Differences between pseudo-collective and pseudo-distributive primes correspond to differences in the verification strategies that could be developed to make an accurate choice.

the collective/distributive ambiguity. Instead, pseudo-primes were associated with distributive/collective conditions through the verification strategies mentioned above: decisions could be based on one (collective) or on more than one (distributive) scale (see Table 3.3).

There were then two types of pseudo-primes. In *pseudo-collective primes*, the sentence involved a singular definite description (e.g. “The book is light”). Correct and foil pictures only differed on one scale (i.e. the scale containing the relevant exemplar). As in collective primes (cf. Experiment 1), participants could make an accurate decision by verifying a single scale. Collective and pseudo-collective primes shared a “check-one-scale” strategy. In contrast, in *pseudo-distributive primes*, the sentence included the focus-sensitive operator *only*, which enriches the meaning of the expression by negating its alternatives. A sentence such as *Only the book is light* implies that nothing but the book is light. Indeed, the foil image for these trials made the sentence false by satisfying the predicate not only for the exemplar but also for its alternatives. Since foil and correct images differed on more than one scale, participants are required to check at least two scales in each image. As a result, distributive and pseudo-distributive primes give rise to a common verification strategy.

If the effect found in Experiment 1 was due to verification strategy priming only, we would expect both pseudo-primes and primes that instantiate the same verification strategy to behave similarly. Instead, semantic priming should lead to an effect for targets specifically after *true* primes, and not after *pseudo*-primes. Then, using pseudo-primes as baselines, the potential difference between primes and pseudo-primes within one condition, say distributive, could

be used as a measure of the priming force of distributive primes, and we could look for asymmetries between distributive and collective priming.

The experimental design was the same as in Experiment 1 except that there was an additional PRE-TARGET TYPE factor (prime, pseudo-prime). The PRIME CONDITION is defined here by the verification strategy (see Table 3.3). We obtained 32 experimental triplets (96 trials) by crossing the factors. A further 64 controls were randomly inserted between triplets.

A group of 55 fresh English speakers was recruited using Amazon Mechanical Turk. After exclusion (cf. Experiment 1), 41 participants were included in the analyses. Responses were analyzed as in Experiment 1.

PRIMING SOURCE	Prime	Pseudo-Prime
Reading Priming	YES	NO
Verification strategy priming	YES	YES

Table 3.3: Items in Experiment 2. Primes and pseudo-primes in Experiment 2 differed on whether they could serve to prime both a verification strategy and a semantic interpretation or only a verification strategy. While pseudo-primes do not force collective or distributive interpretations, they share the verification strategy with the prime from the same condition.

3.3.2 Results and discussion

Figure 3.4a illustrates the proportion of distributive responses after pseudo-primes and primes. First, results from Experiment 1 were replicated: when restricted to targets after *prime* trials (Figure 3.4a, right panel), the proportion of distributive choices is significantly higher after distributive primes than after collective primes ($p < .001$, $\chi^2 = 35.2$).

Moreover, the difference between collective and distributive primes is significantly different from the one between pseudo-primes, as it is revealed by a significant interaction PRIME CONDITION (collective/distributive) \times PRE-TARGET TYPE (prime/pseudo-prime) ($p < .001$, $\chi^2 = 17.1$). This result suggests that the priming effect is not entirely driven by verification strategy priming, or else primes and pseudo-primes would have had the same effect in targets. Therefore, priming of semantic interpretation must be at play.

To assess whether the priming effect was asymmetric, we analyzed the proportion of responses compatible with priming within each condition (see Figure 3.4b); namely, the responses compatible with distributive readings in the distributive condition, and the ones compatible with the collective reading in the collective condition. If, e.g., distributive priming were stronger than collective priming (cf. Maldonado et al 2017), the difference between pseudo-primes and primes in the distributive condition should be bigger than in the collective one. The interaction PRIME CONDITION \times PRE-TARGET TYPE, however, was not significant ($p = .49$, $\chi^2 < 1$), indicating no evidence of an asymmetry in distributive and collective priming.

Finally, a *posthoc* analysis revealed a baseline preference for collective interpretations (main effect of PRIME CONDITION) in the experiment. This goes in the same direction as previous results in the literature (Brasoveanu & Dotlačil, 2015; Frazier, 1999; Syrett & Musolino, 2013).

These results thus suggest that the collective/distributive contrast gives rise to priming of semantic interpretation, independently of verification strategy priming (which may be at play on top of the semantic effect). There was no evidence of an asymmetry of priming from distributive or collective primes. Unlike previous experiments then (cf. Maldonado, Chemla, & Spector, 2017), our findings cannot be straightforwardly explained by saying that priming targets the mechanism responsible for distributive interpretations (i.e. distributivity operator). Instead, some part of the abstract representations underlying *each* of the readings seem to be the locus of the priming effect. The absence of replication of an asymmetry in priming could be

related to the use of different contrasts (cumulative/distributive vs. collective/distributive) or to predicate types (transitive vs. adjectival).

Alternatively, one could propose that the derivation of collective readings also involves some abstract mechanism, analogous to the distributivity operator, which can be primed. Unlike cumulative readings, collective interpretations would not necessarily be considered primitive interpretations. Something along these lines has been proposed by Winter (2001).

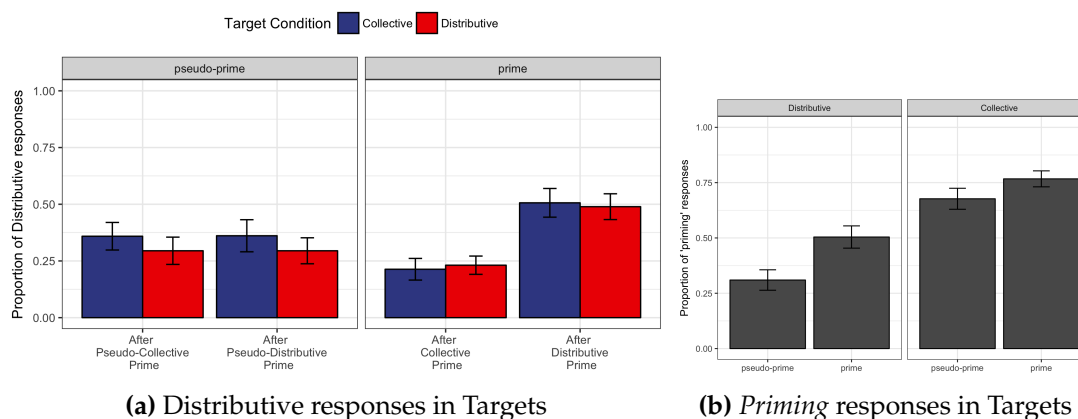


Figure 3.4: Target results in Experiment 2. (a) Proportion of responses compatible with the distributive reading on targets trials. (b) Proportion of responses compatible with *priming* within each prime condition. Responses are coded depending on whether or not they are compatible with distributive readings (left panel) or collective readings (right panel). Target condition is aggregated for simplification.

3.4 Conclusions

Distributive readings are often assumed to arise from applying an invisible distributivity operator D . This proposal has led both theoretical and psycholinguistic studies to try to confirm or disconfirm the existence of a mechanism responsible for distributive interpretations. Such studies have mostly been based on distributivity associated with transitive predication. Instead, we proposed to investigate the specific case of adjectival predicates, which have remained rather unexplored in the literature (but see [Scontras & Goodman, 2017](#)). Priming results revealed a substantial difference between distributive and collective interpretations and did so, for the first time, independently of considerations of visual priming, object covariation, logical strength and verification strategies.

Under the view that the priming effect is due to the presence of the distributivity operator D in both primes and targets, one may expect to find that the priming force of distributive primes would be higher than that of primes with an *absence* of D . Evidence for this view had been found in [Maldonado, Chemla, & Spector, 2017](#), but were not replicated here. Although the current study has relevant differences (e.g., it contrasts distributivity with collectivity rather than with cumulativity), this discrepancy calls for further investigation which should inform us both about distributivity and about the mechanisms by which priming operates in semantics ([Maldonado, Spector, & Chemla, 2017](#)).

3.5 Supplementary materials

Control trials Four types of **control trials** (see Figure 3.5) were included in the experiment. *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. *Strong-Distributive* and *Strong-Collective* controls 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 correspond to a scenario than makes both readings true (strong picture). In the same way as 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 collective and distributive readings.

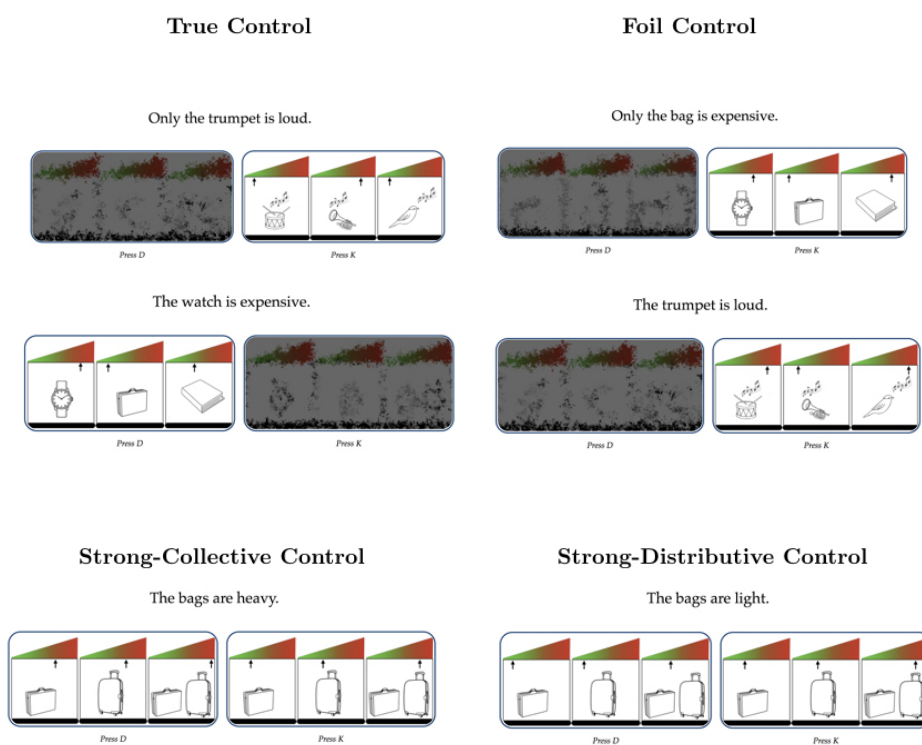


Figure 3.5: Illustration control trials

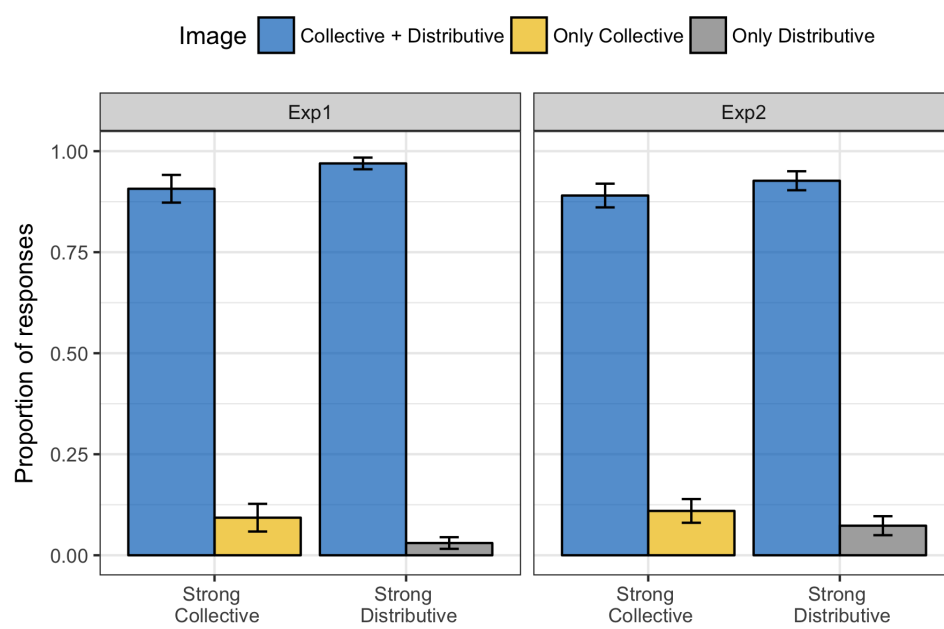


Figure 3.6: Control results for both experiments.

4 | Plural ambiguities through priming: A brief discussion

In Chapters 2 and 3, I presented two priming studies that seek to identify the abstract mechanisms responsible for plural ambiguities. Both of these studies suggest that non-distributive and distributive readings are associated with different semantic representations, which can be the locus of priming.

In this chapter, my goal is to compare the two studies, discussing alternative explanations of the results. First, I explore in more detail the semantic contrasts used in each study, and discuss potential differences between them. As a second step, I comment on a puzzle that emerges from differences in results between the studies, providing additional suggestions to understand this data.

4.1 A closer look at the non-distributive/distributive contrast

4.1.1 Do cumulative and collective interpretations arise from the same representation?

So far, we have assumed that the contrasts tested in Chapters 2 and 3 were roughly two instances of the same non-distributive/distributive ambiguity. That is, despite minor differences, cumulative and collective readings were taken to be two readings of a single non-distributive LF.

The predicates used in our studies, however, do not allow making a three-way distinction between collective, cumulative and distributive readings. For example, it is unclear that the sentence ‘Two squares are connected to two circles’ (see Chapter 2) can have a collective understanding such that a group of two squares is jointly connected to a group of two circles, without each square being individually connected to one of the circles. Unlike collective readings, cumulative readings are typically taken to involve two plural entities A and B in a “cross-product” like relation (Champollion, to appear; Scha, 1981, 1984), such that each member of A is in a relation with at least one member of B, and each member of B is in a relation with at least one member of A. Consequently, it is impossible to obtain cumulative readings of gradable adjectives –and more generally, of any intransitive predicate.

Nevertheless, a distinction between collective, cumulative and distributive readings can be instantiated within one sentence. Consider, for instance, sentence (1):

- (1) Two students invited two linguists to the party.
 - a. Two students each invited two different linguists to the party. Four linguists were invited in total.
 - b. Two students each invited one linguist to the party. Two linguists were invited in total.
 - c. Two students together invited two linguists to the party, without each separately doing so (i.e. joint invitation).

Besides having a distributive reading compatible with the covariation scenario in (1a), sentence (1) has a cumulative reading (compatible with (1b)) and a collective reading (compatible with (1c)).

Debates have centered around the question of whether or not these collective and cumulative interpretations reflect the existence of two distinct semantic representations or LFs. The picture presented in Chapter 1 suggests that these two readings arise from a single, underspecified, logical form (Link, 1987; Roberts, 1987), obtained as soon as the predicate is combined with the plural subject:

- (2) $\exists x.\exists y. |x| = 2 \wedge \text{student}(x) \wedge |y| = 2 \wedge \text{linguist}(y) \wedge \text{invite}(x, y)$
 There is a plurality made up of two students, call it x , and a plurality made up of two linguists, call it y , such that x invited y to the party.

Under the *lexical cumulativity hypothesis*, lexical predicates are closed under mereological sum. If the predicate is true of individuals, it is also true of the plurality made up of those individuals (Kratzer, 2007; Krifka, 1992; Scha, 1981). Moreover, if the predicate is “mixed”, it can be true *only* of pluralities. The relation *invite*(x, y) –in its plural form– is therefore true in both scenarios (b) and (c), since it does not matter whether the individuals that make up x are collaborating on the inviting event or not.

Alternatively, some authors have suggested that –at least some– collective readings are inherently different from cumulative ones, revealing a true ambiguity (Champollion, 2010; Landman, 1996, 2000; Winter, 2001). Most of these approaches account for cumulative readings by closing the predicate under sum, but put forward an additional mechanism for collective interpretations. The basic idea is that collectivity involves a predicate that applies to a plural entity as a whole, beyond its individual members, and therefore the plurality should not be modelled as a sum. For instance, collectivity is often derived by applying a group formation operator –the \uparrow -operator in (3)– to the plural subject (Champollion, 2010; Landman, 2000). The \uparrow -operator introduces a distinction between the sum $y \oplus z$, whose proper parts are y and z , and the group that corresponds to this sum ($\uparrow(y \oplus z)$), which is taken to be a mereological impure atom that has no proper parts (no parthood relation with its members) (Champollion & Krifka, 2014).¹

Simplifying drastically, the collective reading of (1) on this view would be (4).²

- (3) Group formation (primitive function) : $x = \uparrow(y \oplus z)$, where x is an impure atom.
 (4) $\exists x.\exists y. |x| = 2 \wedge \text{student}(x) \wedge z = \uparrow(x) \wedge |y| = 2 \wedge \text{linguist}(y) \wedge \text{invite}(z, y)$
 There is a group of two students, call it z , and a plurality made up of two linguists, call it y , such that z invited y to the party.

Cumulative readings would then be obtained as soon as the predicate is applied to the plural subject, whereas both collective and distributive interpretations would be derived, even though in different ways –by applying an operator either to the predicate or to the plural subject. For our purposes, this distinction is relevant as long as it might constitute an alternative explanation of the discrepancies in results between our priming experiments. In particular, one could conceive the possibility that priming can only target the presence of operators but not their absence. That is, we could prime the distributivity operator (D) as well as the group formation operator (\uparrow). This would explain why we obtain an asymmetry in strength between

¹Impure atoms were originally modelled to account for group nouns like ‘committee’ or ‘gang’ (Landman, 2000). Furthermore, the existence of a group formation operator also serves to explain sentences such as ‘The linguists and the philosophers each won an award’, where the quantificational ‘each’ distributes over each of the impure atoms (Landman, 1996, 2000). The reader is referred to Schwarzschild (1996) for an account of these facts in terms of covers.

²Note that once impure atoms and the group formation operator are introduced into the system, one might need to allow predicates to range over these impure atoms as well as over pure atoms and sums. See Champollion (2010) for discussion.

cumulative and distributive priming (Chapter 2), but not between collective and distributive priming (Chapter 3).

4.1.2 Sources of distributivity: transitive predicates *versus* gradable adjectives

An additional question is whether all distributive readings correspond to the *same* type of distributive representation. That is, whether distributive readings of transitive and adjectival predicates underlie the same derivational mechanisms.

We adopted the view that predicates that give rise to a non-distributive/distributive distinction are “mixed” predicates, in that they can range over both atoms and pluralities (Landman, 1996; Link, 1987, see Chapters 1 and 3).³ As observed, distributive entailments are not mandatory for these “mixed” predicates. For instance, (5a) does not necessarily entail (5b).

- (5) a. Amir and Milica carried a box.
b. Amir carried a box and Milica carried a box.

Distributive inferences for sentences involving “mixed” predicates were assumed to arise by inserting an additional distributivity operator *D*, which applies the predicate to each atomic member of the subject (Champollion, to appear; Link, 1987). That is, the LF [Amir and Milica [D[carried a box]]] does entail (5b).

The behaviour of “mixed” predicates contrasts with the one of predicates such as ‘smile’, for which the entailment is obligatory:

- (6) Amir and Milica smiled. \Rightarrow Amir smiled and Milica smiled.

Here, distributivity is taken to arise directly from lexical semantics, i.e. from properties of the specific predicate (Champollion, 2010, to appear; Landman, 1996, 2000; Lasersohn, 1989; Link, 1983; Winter, 2001, see Chapter 1). For instance, predicates like ‘smile’ might include a distributive requirement as part of their lexical entry (as a meaning postulate) or they might only contain atoms in their original extension.⁴

The boundary between these two sources of distributivity (from an operator and from lexical semantics) is often controversial. In particular, some approaches have suggested that the distributivity operator is only required to derive distributive interpretations compatible with covariation scenarios; namely, to distribute objects over members of the subject. In contrast, distributive inferences that do not involve covariation—typically of intransitive predicates—might all be obtained by means of the same mechanism, possibly from lexical semantics (Vries, 2015; Winter, 2001, among others). Under this view, the existence of distributive and non-distributive interpretations for intransitive predicates is a matter of underspecification rather than a true ambiguity.

Along these lines, recent experimental evidence has suggested a difference in processing between distributivity arising from transitive and intransitive predicates. Brasoveanu and Dotlačil (2015) compared the processing pattern of sentences such as ‘Amir and Milica won’ to their transitive alternatives (e.g. ‘Amir and Milica won an award’) during a self-paced reading

³I focus here on the collective/distributive contrast for transitive and intransitive predicates, but similar observations hold for the cumulative/distributive distinction discussed in Chapter 2.

⁴Instead of making stipulations about specific lexical items, some approaches have proposed that distributive inferences arise from general reasoning about what predicates could possibly mean (Glass, n.d.; Vries, 2015; Winter & Scha, 2015). For example, by world knowledge, we know that smiling is an action performed by individual faces, therefore whenever we are told that two people smiled, we infer that each of them did. However, given that distributivity does not arise from the predicate semantics *per se*, under the right circumstances, collective understandings are expected to arise (e.g. ‘Milica’s lips smiled’, Winter & Scha, 2015). Note that one consequence of these accounts is that all predicates are fully underspecified for distributivity or collectivity. Different readings arise as pragmatic inferences.

task.⁵ Distributive readings were found to carry an additional processing cost (slow-down in reading) for transitive, but not for intransitive sentences. Brasoveanu and Dotlačil then suggested that two different types of distributivity are at play in these two sentences.

As observed in Chapter 3, these findings could also be interpreted by saying that additional processing cost in transitive sentences is the result of covariation of objects, and it might not be related with distributivity *per se*. Still, one might wonder whether distributive readings of adjectival predicates (e.g. ‘heavy’, ‘light’) can be directly derived from their lexical meaning. If is the case, our two priming studies might be targeting two different phenomena.

It is easy to note that assuming lexical distributivity would lead to the wrong results for predicates such as ‘heavy’. The sentence ‘The bags are heavy’ can be true in a scenario where each of the bags is not heavy (i.e. collective scenario), indicating that the predicate ‘heavy’ can be true of a plurality without being true of its members. Moreover, one should note that ‘heavy’ is closed under sum due to its lexical semantics: given our knowledge about weight, if two objects a and b are heavy, so is $a \oplus b$ ($\llbracket \text{heavy} \rrbracket = \llbracket * \text{heavy} \rrbracket$). To account for collective readings, the predicate has to be “mixed”: it should be able to range over pluralities without ranging over the members of the plurality. Distributive inferences can therefore only be obtained by applying some additional mechanism.

On the other hand, adjectives such as ‘light’ do behave like other lexically distributive predicates (e.g. ‘smile’) in that they always give rise to distributive entailments: whenever the predicate is true of a plurality, it is also true of its constituents. This is reflected in the entailment pattern mentioned in Chapter 3. While this behaviour has led some authors to suggest that ‘light’ is lexically distributive (Glass, 2018), the account for other lexically distributive predicates cannot be easily extended to ‘light’.

Assume that the predicate ‘light’ is lexically distributive and closed under sum. Its denotation should always contain atoms as well as their sums. Like for ‘smile’, the following entailment should thus hold: $\text{light}(a) \wedge \text{light}(b) \Leftrightarrow \text{light}(a \oplus b)$. Now, consider the sentence ‘The bags are light’. In an scenario where every bag is individually light, this sentence should be automatically judged true. This, however, does not seem to be the case. The sentence ‘The bags are light’ seems to have a collective reading, which will only be true if the *total* weight of bags is less than a certain value. The existence of two different readings that have different truth-conditions suggests that ‘light’ gives rise to an ambiguity in a way that ‘smile’ does not, even if the distributive entailment does indeed hold. To derive this ambiguity, an additional mechanism (such as the D operator) is required.

As far as I can see, the facts presented above suggest that the general picture in Chapter 3 is on the right track, at least regarding gradable adjectives: predicates such as ‘light’ and ‘heavy’ give rise to a true ambiguity between collective and distributive readings. The question of whether the semantic mechanism behind the non-distributive/distributive distinction is *exactly* the same for transitive and intransitive VPs, however, has to be addressed in future research⁶.

Before concluding this section, I would like to make a minor remark about how the properties of ‘heavy’ and ‘light’ fit in the general categorization of gradable adjectives. Most gradable adjectives do not give rise to collective understandings when combined with plural subjects. Following Glass (2018), a distinction can be made between adjectives that are *truly* distributive, in that there is no possible collective understanding (e.g. ‘old’, ‘full’, ‘new’), and those for which one could imagine a collective reading, but strongly favor a distributive interpretation (e.g. ‘big’, ‘tall’). The latter are the so-called ‘stubbornly distributive’ predicates (Quine, Churchland, & Føllesdal, 1960; Schwarzschild, 2011, a.m.o.).

⁵These two sentences can be interpreted either distributively —if each of Amir and Milica won (an award)— or collectively —if they won as a team.

⁶Arguably, this question could be addressed by testing the existence of priming effects *between* predicates (transitive and adjectival).

In a recent study, [Scontras and Goodman \(2017\)](#) suggested that these ‘stubbornly distributive’ predicates can get collective understandings with the right pragmatic support. Indeed, the dispreference for collective interpretations seems to correlate with how predictable these understandings are in a given context. For instance, joint height is typically less predictable than joint weight, explaining why collective readings tend to be less available for ‘tall’ than for ‘heavy’. Interestingly, under [Scontras and Goodman’s](#) account, priming could potentially target distributive and collective interpretations of ‘stubbornly distributive’ predicates whenever the collective interpretation is made predictable in the context.

In light of these observations, it is surprising that the ambiguous sentences in Chapter 3 display a preference for collective understandings (see Figure 3.2). Given that distributive readings are always available for gradable adjectives, one might expect people to draw an analogy and assign distributive interpretations by default. Instead, our data suggest that if the collective interpretation of the predicate is accessible, it will be preferred across the board.

4.1.3 Different accounts of the distributive/non-distributive contrast (added on April 23rd)

We have been working on the assumption that distributive and non-distributive readings correspond to two different LFs. That is, plural ambiguous sentences were taken to be a case of *structural ambiguity*. Our results were therefore interpreted as priming of the mechanisms responsible for the derivation of each of these LFs. There are, however, alternative accounts of plural ambiguities. What do our results suggest about the choice between these approaches? In other words, can our findings distinguish between different theories? I will explore [Schwarzschild’s \(1996\)](#) and [Kratzer’s \(2007\)](#) views (see Chapter 1 for a brief introduction).

On [Schwarzschild’s](#) view, plural ambiguous sentences are taken to be instances of context dependence, rather than cases of *structural* ambiguities: a single LF or structure can give rise to different readings depending on what kind of cover is contextually given. That is, the LF contains a contextual parameter which needs to be fixed in order to obtain the sentence truth conditions. This is not radically different from the structural ambiguity view: presumably, the selection between two LFs is also driven by contextual factors.

While priming effects are traditionally thought to rely on structural differences, priming may also target the underlying factors that drive the specification of the reading, including contextual factors, such as the choice of the cover. In our studies, for instance, distributive primes make relevant a cover where the plurality is divided into individuals. Conversely, collective primes make relevant a cover where the plurality is considered as a whole. Priming might target these different covers, explaining our results. Given that our experiments use different predicates and objects between primes and targets, one needs to posit that priming taps onto a rather abstract property of covers.

To test this possibility, one could try to manipulate the choice of cover independently from the distributive/non-distributive resolution, and see whether this also creates priming effects. Given the similarity between specific covers and verification strategies, one might think that the pseudo-primes (i.e. baselines) in Chapter 3 would serve to address this goal. Indeed, pseudo-primes forced similar verification strategies as primes (*check one scale* vs. *check two scales* strategies), without displaying the relevant ambiguity. However, baseline sentences did not involve plural elements (e.g. “The book is heavy”): since no cell contained multiple elements, there is no cover of a plurality which could be primed. Hence, at this point, [Schwarzschild’s](#) account is compatible with our results.

What about [Kratzer’s](#) approach to the non-distributive/distributive contrast? By requiring the application of the star-operator to the sister of the plural subject, Kratzer derives an LF whose truth conditions are compatible with different scenarios (i.e. non-distributive and distributive ones). On this view, there is a single weak reading, so there is no ambiguity to be resolved. In fact, there is no real intrinsic distinction between a distributive and a non-distributive reading.

As a result, an effect of priming would have to be explained by heterogeneous factors. For example, in our studies, different primes would be forcing the same LF, but just differing on the scenario that makes true this reading (in this case, the type of picture). The existence of an effect of primes in targets could only be interpreted as visual priming. As observed in Chapters 2 and 3, effects due to visual priming were controlled for, and ruled out as possible explanations for our findings. Consequently, our results seem to argue against at least this simple interpretation of Kratzer's view.⁷

4.2 A note on preference and asymmetry differences

Our priming results differ on whether or not they display an asymmetry in the strength of the effect: while, in Chapter 2, distributive primes give rise to stronger priming effects than cumulative ones, in Chapter 3 there is no evidence of an asymmetry between distributive and collective priming. Priming strength—the magnitude of the increase in probability of assigning the same interpretation in targets as in primes—can be influenced by at least two different factors.⁸

First, one could imagine that asymmetries in strength result from inherent differences in the locus of the priming effect (Branigan & Pickering, 2017; Chemla & Bott, 2015). For instance, priming effects might reveal that representations in prime and target are equivalent (*representation priming*), or that the same operation applies equally in both cases (*operation priming*). If priming targets representations themselves, no difference in strength is *a priori* expected: different representations should be equally primed. Instead, if operation priming is at play, one would predict priming effects to arise only after primes that instantiate the operation, leading to potential asymmetries.

On the other hand, the degree of preference for each reading might influence priming strength, originating inverse preference effects, whereby less common structures give rise to stronger priming than more frequent ones (see Chapter 2, Branigan & Messenger, 2016; Ferreira & Bock, 2006; Jaeger & Snider, 2013; Myslin & Levy, 2016).

On error-based accounts of sentence comprehension (Chang et al., 2006; Fine et al., 2013; Jaeger & Snider, 2013; Kuperberg & Jaeger, 2016; Levy, 2008), inverse preference effects are taken to be a consequence of adaptation, with the goal of minimizing future prediction error. The idea is that comprehenders have specific expectations about what sentences should mean. When an unexpected interpretation (dispreferred or infrequent) is forced in primes, they will experience a prediction error and adjust their expectations accordingly, increasing the probability of assigning this dispreferred reading to the subsequent target. Priming effects are just an indication of this prediction adjustment. Therefore, the strength of the priming effect is expected to be positively correlated with the prediction error (i.e. the larger the error, the stronger the priming). Importantly, to assess whether inverse preference effects are at play, one needs to have a measure of comprehenders' expectations, which seem to be sensitive to both prior and recent experience within an experiment.

In what follows, I explore how these factors might or might not explain the difference in results between Chapter 2 and Chapter 3. I consider four possible alternatives: (a) operation priming; (b) inverse preference effects relative to preference rates observed in the experiment;

⁷A slightly modified version of this proposal is one where applying the star operator is optional rather than mandatory. As a result, there would be two possible LFs: a 'starred-LF', compatible with both readings, and a basic LF, obtained by lexical cumulativity as soon as the predicate is combined with the plural subject. The later LF is only compatible with non-distributive readings. Distributive primes would prime the starred LF, whereas cumulative primes would presumably prime the more basic LF. Our results are compatible with such a view, which in practice is very close to the standard structural ambiguity view: while in one case there is a one-to-one mapping between structures and readings, in the other there is a one-to-many mapping.

⁸Nothing prevents multiple factors to be at play on top of each other, resulting in the different priming strength patterns.

(c) inverse preference effects relative to prior preference rates (and their connection with logical strength relations); and (d) ceiling effects combined with operation priming. Note that all these alternative explanations are speculative: at this moment, there is no clear way of teasing these options apart.

4.2.1 Priming the presence of an operator

The results reported in Chapter 2 were interpreted as an instance of *operation priming*. The asymmetry in priming strength (i.e. stronger distributive priming)⁹ was accounted for by saying that priming was partially targeting the mechanism responsible for distributive interpretations. We thus concluded that the *presence* of a distributivity operator or mechanism could be specifically primed, but not its absence.

An explanation in terms of *operation priming* cannot be easily extended to the results in Chapter 3, where there was no evidence for an asymmetry in priming strength: collective and distributive priming were observable to the same extent. Instead, priming might be targeting semantic representations themselves.

Alternatively, as sketched in Section 4.1.1, a semantic distinction between cumulative and collective interpretations might be at the core of the strength difference. If both collective and distributive readings are derived, one might expect priming to arise for both of these readings to a similar degree: priming would be targeting the specific semantic mechanisms involved in the derivation (e.g. distributivity/collectivizer operators). This would not be expected for more basic cumulative readings.

4.2.2 Inverse preference effects relative to baseline preference rates

Our first set of results (Chapter 2) revealed a preference for *distributive* interpretations within the experiment (~ 60% mean percentage of distributive choices in baseline targets). This result seems to conflict with previous findings, which have mainly found a strong preference for non-distributive/collective readings (Brasoveanu & Dotlačil, 2015; Brooks & Braine, 1996; Dotlačil, 2010; Frazier, Pacht, & Rayner, 1999; Syrett & Musolino, 2013, among others). We have, however, claimed that this incompatibility is only superficial (see Chapter 2).

The probability of assigning one particular interpretation to an ambiguous sentence seems to be dependent not only on prior experience but also on the experimental set-up, the use of specific lexical items and recent experience, as suggested by the mere existence of priming and adaptation effects (Degen, 2013; Fine et al., 2013; Jaeger & Snider, 2013). In the context of our experiment, participants might come to have higher expectations for *a priori* dispreferred interpretations (i.e. distributive readings), resulting in the attested baseline rates.

While in Chapter 2 we tested the cumulative/distributive distinction, in Chapter 3 we investigated the collective/distributive one. The findings reported in Chapter 3 indicate that this difference in contrasts might play a role: in line with previous research, we found a baseline preference for *collective* interpretations (i.e. the proportion of distributive responses is below 50% for both baseline conditions). Note, however, that, in both experiments, baseline rates are roughly balanced; namely, there is no strong dispreference for any of the sentence interpretations.

Despite being unstable across experiments, baseline preference rates *within* a priming experiment can be used to estimate comprehenders' expectations and to compute the prediction error triggered by primes. This way, one can assess the presence of an inverse preference effect.

⁹While our results suggest an asymmetry in priming strength, they do not indicate that cumulative priming does not exist: cumulative priming is observable as an accumulative effect, which might cause a decrease in the overall amount of distributive choices. I will not explore this effect here. One should, however, note that this effect, which can be seen as long-lasting priming or adaptation, might interact with trial-to-trial priming discussed in this section.

Our results argue against such an effect: while, in Chapter 2, distributive interpretations are slightly preferred and more strongly primed, in Chapter 3 both distributive and collective readings are primed to the same extent, despite a dispreference for distributive interpretations. As a result, differences in priming strength between studies cannot be explained by an inverse preference effect, at least when this effect is computed for the baseline rates observed in the experiment.

4.2.3 Inverse preference effects relative to prior expectations

Alternatively, one could take into consideration the influence that *prior* expectations have on priming strength. Our findings could be analysed as the result of an inverse preference effect arising from differences in prior preference rates —instead of from differences in baseline preference rates within the experiment. This hypothesis relies on two main assumptions.

First, prior experience should be taken to affect prediction error in a way that is not reflected by baseline preference rates within the experiment (some indirect evidence for prior experience influencing error adjustment can be found in Jaeger & Snider, 2013). That is, baseline rates can no longer be taken as a good measure of comprehenders' expectations, nor be used for estimating the prediction error.

Second, the degree to which one interpretation is preferred over another is assumed to be dependent not only on the type of reading (whether the reading is distributive, cumulative or collective), but also on its logical strength (whether or not the reading is entailed by its alternative). Comprehenders' expectations are specified with respect to both of these factors in a non-additive way. For instance, a dispreference for distributive over non-distributive readings might only arise when these interpretations are logically independent. Instead, weak interpretations might all be equally dispreferred, independently of whether they correspond to the distributive or to the non-distributive reading of the sentence.

In Chapter 2, we compare priming of two logically independent readings (i.e. there are no weak interpretations). If prior experience favours cumulative readings over distributive readings, inverse preference effects based on prior expectations could directly explain the asymmetry in priming strength, without appealing to an explanation in terms of priming of the distributivity operator.

Instead, in Chapter 3, prime trials always present a choice between a foil and a *weak* picture, forcing the access to the *weak* interpretation of the sentence, collective or distributive depending on the case (see Table 3.1). These *weak* readings —both collective and distributive— also happen to be drastically dispreferred with respect to strong interpretations (see Figure 3.6). After accessing a strongly dispreferred reading in primes, a boost of priming strength is expected in targets. Since both prime conditions force weak interpretations, and all weak readings are *a priori* equally dispreferred, we can explain why no difference in priming strength is visible in Chapter 3. I will call this general boost in priming strength a “general inverse preference effect”.

One could then argue that if, as proposed, priming strength depends on logical strength, we would have expected priming to arise only for targets instantiating weak interpretations. This, however, was not the case. Despite using weak dispreferred interpretations in primes, priming effects in Chapter 3 were shown to be independent of logical strength: *weak* distributive readings can serve to equally prime both *weak* and *strong* distributive interpretations in targets. In order to maintain the view above, one should posit that, while the prediction error is affected by general expectations, the adjustment triggered by the prediction error (priming effect) is itself reading specific: for instance, forcing weak distributive readings would result in a boost of priming of all kinds of distributive interpretations (strong and weak; *mutatis mutandis* for collective readings).

4.2.4 A hybrid explanation: ceiling effects

Finally, we could keep an explanation in terms of operation priming for the results in Chapter 2 (i.e. priming the distributive mechanism), and take the lack of asymmetry between distributive and collective priming in Chapter 3 as being the result of a *ceiling effect*. A ceiling effect would be attested in priming results whenever the impact of primes in targets reaches the *maximum* degree to which primes can possibly influence targets.

I would like to raise the possibility that a ceiling effect might be at play in Chapter 3, whereby the upper limit for priming effects has been reached and no room is left for additional, *asymmetric* priming. Indeed, the impact of primes in target responses—quantified here as the difference between mean target responses after each prime—is larger in Chapter 3 (Exp1: Δ distributive-responses = .25, $d = .44$) than in Chapter 2 (Exp1: Δ distributive-responses = .1, $d = .2$).¹⁰ If the priming effects reported in Chapter 3 are actually occupying the whole ‘priming space’, extra priming of the mechanism responsible for distributive readings would no longer be detected.

One might then ask why a ceiling effect is only visible in Chapter 3. Arguably, the fact that priming effects are overall stronger in Chapter 3 than in Chapter 2 may be due to the aforementioned “general inverse preference effect”: forcing strongly dispreferred, weak interpretations results in a boost of the priming effect across the board (i.e. for both prime conditions). In contrast, the two readings in Chapter 2 are not particularly imbalanced in terms of preference (as indicated by baseline rates), so no extra boost is expected here.

¹⁰The Δ value indicates the difference in mean distributive responses for targets after non-distributive primes and after distributive primes; Cohen’s d corresponds to this difference expressed in standard deviation units, which is a common effect size measure.

Part II

Beyond plural ambiguities

Can the questions and methodology developed in Part I be applied to other semantic and pragmatic phenomena? In Chapter 5, I discuss the use of priming as tool in semantics and pragmatics. In Chapter 6, I present a study where a priming method is use to investigate the semantic representations and operations underlying scopally ambiguous sentences. In Chapter 7, I show how the mechanisms behind plural and scope ambiguities can interact, generating inverse-scope distributive interpretations. These readings usually go unnoticed by simple introspection but can be detected through experimental manipulations in acceptability judgments.

5 | Priming methods in semantics and pragmatics*

Maldonado, M, Spector, B. & Chemla, E. [Priming methods in semantics and pragmatics](#). *Behavioral and Brain Sciences*. 2017. Commentary on: Branigan, H., & Pickering, M. (2016). [An experimental approach to linguistic representation](#). *Behavioral and Brain Sciences*. 2016.

Abstract

Structural priming is a powerful method to inform linguistic theories. We argue that this method extends nicely beyond syntax, to theories of meaning. However, priming should still be seen as only one of the tools available for linguistic data collection. Specifically, because priming can occur at different, potentially conflicting levels, it cannot detect every aspect of linguistic representations.

[Branigan and Pickering](#) (henceforth, B&P) argue that structural priming is a powerful method to inform theories of linguistic structure, and they even suggest that it could supersede other methods such as acceptability judgments. We will argue that the method extends nicely beyond syntax, to theories of meaning, where priming can serve to reveal abstract interpretive operations. In doing so, however, we will see why structural priming should still be seen as one among many of the tools available for linguistic data collection. In particular, because priming can occur at different, potentially conflicting levels, it cannot detect every aspect of linguistic representations.

The primary data used in formal semantics/pragmatics are truth-value inferential judgments. These methods document the result of interpretive processes: what a sentence ends up meaning. Priming methods can be useful here just like in syntax, where, schematically, acceptability judgments target the output of a cognitive process, while priming data may offer a window into some aspects of this process. In formal semantics and pragmatics, the relevant elementary interpretive processes are often abstract “invisible” operations. Here are two prime examples:

- A silent distributivity operator whose meaning is akin to *each* has been postulated to explain why sentences involving more than one plural expression, such as *Two boys read three books*, have both a cumulative interpretation (i.e. two boys read three books *in total*) and a distributive interpretation (i.e. two boys read three books *each*) —see [Champollion to appear](#), for a survey.
- Sentences such as *Some of the students came* tend to acquire a strengthened meaning amounting to *some but not all of the students came* (scalar implicature). On some accounts, this strengthening is a pragmatic process, while on others, it is due to the presence of a covert exhaustivity operator. In both approaches, however, the very same mechanism is

We wish to thank Tal Linzen for useful comments and discussion. The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007–2013)/ERC Grant Agreement n.313610 and from the Agence Nationale de la Recherche (Grants ANR-10-LABX-0087 IEC, ANR-10-IDEX-0001-02 PSL and ANR-14-CE30- 0010-01 TriLogMean).

responsible for the strengthening of *some* into *some but not all*, *may* into *may but does not have to*, *three* into *exactly three* (though this last case is more controversial).

Distributivity and exhaustivity operators are often thought as covert linguistic operators, which are part of syntactic representations (Chierchia, Fox, & Spector, 2012; Link, 1983, 1987). Alternatively, exhaustivity operators can also be seen as pre-compiled proxies for late, post-compositional pragmatic processes (Spector, 2007b; Van Rooij & Schulz, 2004). In either case, the possibility of priming exhaustive or distributive interpretations across sentences featuring different trigger words would confirm the posited abstract mechanisms, beyond their mere interpretive effects. Indeed, recent studies using typical priming paradigms (e.g., Raffray & Pickering, 2010) have provided evidence in favor of an abstract mechanism for distributivity (Maldonado, Chemla, & Spector, 2017) and exhaustivity (see L. Bott & Chemla, 2016; Chemla & Bott, 2014; Rees & Bott, 2018: schematically, the presence of a distributivity/exhaustivity operator in a sentence primes the presence of this operator in a subsequent *distinct* sentence, showing that the cognitive representations associated with both sentences share an abstract property.

Because priming can occur at different levels (e.g., syntax and semantics), it is tempting to assume that *every* aspect of linguistic representations can be primed. From there, B&P suggest that we could use the absence of priming effects triggered by certain hypothetical operations as an argument against theories which posit such operations. But this is too radical, as we will briefly argue, focusing on scope assignment in doubly-quantified sentences (Branigan & Pickering, 2017; Chemla & Bott, 2015; Feiman & Snedeker, 2016; Raffray & Pickering, 2010). First, a specific operation can be well-motivated independently of priming. Second, the absence of a particular priming effect may be due to a conflict with another potential source of priming.

The Quantifier Raising (QR) operation has been hypothesized to explain cases of mismatch between the surface ordering of quantifiers and their relative semantic scope: the sentence *A girl invited every boy* can receive the interpretation *For every boy x , a girl y invited x* , where the universal quantifier takes scope above the existential, reversing the surface ordering at the interpretive level. Importantly, QR is not the only possible mechanism to derive inverse scope (through movement). Crucially, all frameworks, including the Parallel Architecture view (Culicover & Jackendoff, 2005) endorsed by B&P, need to characterize the mapping between syntax and semantics, and thus need *some* mechanism to account for inverse scope interpretations.

Since semantic operations can be primed (cf. above), the inverse-scope operation (whatever it is, under any account) might in principle be primed. As observed in the target article, recent studies (Chemla & Bott, 2015; Raffray & Pickering, 2010) did not observe priming of the inverse-scope operation. Instead, these studies revealed priming of the scopal relation itself: a sentence interpreted with a universal taking scope over an existential quantifier would prime a similar interpretation of a subsequent sentence, with a universal taking scope over an existential quantifier, whether or not these interpretations require inverse-scope of the prime sentence and/or of the target. Priming of the inverse scope operation was not found, but this potential priming effect was pushing in a direction opposite to that of priming of relative scope. The only conclusion we can draw then is that the latter is stronger than the former, and certainly not that an inverse scope mechanism does not exist (especially given that such a mechanism is necessary in any framework to account for interpretive judgments).

Priming can be used to reveal the existence of a (primable) aspect of linguistic representations, in syntax and in semantics as we have shown, but not so much to argue *against* the existence of a (less primable or potentially non-primable) feature. Linguistic theories should (a) represent the primable features, and (b) provide the means to distinguish between more or less primable features. Although (a) is consistent with B&P, one may understand B&P as implying that all aspects of linguistic representations can in principle be equally primed, thus rejecting (b). Such a radical view might lead to an unwarranted bias in favor of less expressive theoretical frameworks, by allowing researchers to ignore all aspects of linguistic phenomenology that are

not detectable in priming experiments.

6 | Priming scope ambiguities

Feiman, R., Maldonado, M., & Snedeker, J. Priming quantifier scope: Reexamining the evidence against scope inversion. Submitted.

Abstract

In a study of quantifier-scope priming, [Chemla and Bott \(2015\)](#) found evidence suggesting that representations of quantifiers' relative scope can be primed, but a scope inversion operation cannot. We identify a confound in their materials. In Experiment 1, we replicate their finding with this confound intact. In Experiment 2, we remove the confound and find that all priming disappears. Reviewing the prior literature, we conclude that other priming studies suggest the inversion operation may be primed although only within-quantifiers.

Note: This joint paper is currently under revision by one of my co-authors. I am presenting here a slightly modified version of the paper, based on the original submission.

6.1 Introduction

When we think, we systematically combine concepts to create complex thoughts. When we speak, we systematically combine words to create complex sentences. Since words express concepts and sentences express thoughts, one simple hypothesis is that the syntactic rules for combining words correspond one-to-one to the semantic rules for combining concepts, so that each natural language sentence expresses a unique thought ([Frege, 1891](#); [Montague, 1970](#)). Sentences that are semantically ambiguous—cases where the same sentence, under a single syntactic and lexical analysis, can express two different thoughts—pose a significant challenge to this hypothesis. They are therefore an excellent tool for studying how the rules of syntax and the rules of semantics might differ. One systematic type of semantic ambiguity arises when there are two quantifiers in a clause. For example:

- (1) There is a circle above every star.

This sentence could mean either that there is a single circle, perched above all of the stars, or that, for each individual star, there is some circle above it, potentially a different one in each case (see [Figure 6.1](#)). Since a speaker surely knows which of these meanings they want to convey, they must be forming a representation that is unambiguous, unlike the English sentence. To account for this, linguistic theories (e.g. [Heim & Kratzer, 1998](#); [Hornstein, 1984](#); [May, 1985](#)) posit a level of representation distinct from the surface form of the sentence, typically called Logical Form (LF). LF representations are explicit about the semantic relations between words in a sentence. The two LFs corresponding to each meaning of (1) can be glossed in predicate logic as:

- (2) a. $\forall x[\text{star}(x) \rightarrow \exists y[\text{circle}(y) \wedge \text{above}(y, x)]]$
For every x such that x is a star, then there exists a y , such that y is a circle and y is above x

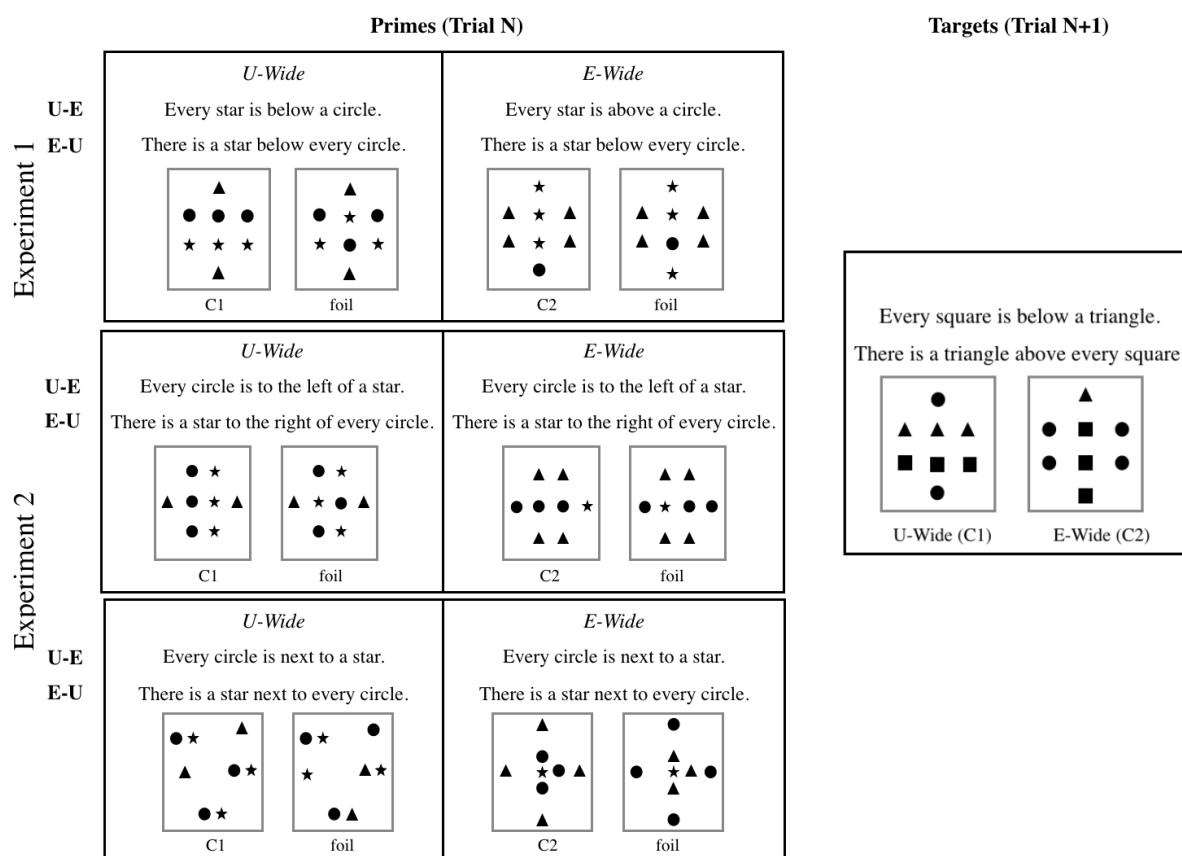


Figure 6.1: Sample prime and target trials in Experiment 1 (upper) and Experiment 2 (lower). Experiment 1 has the predicates ABOVE and BELOW, replicating Chemla and Bott (2015). Experiment 2 used two types of predicates in the primes: LEFT/RIGHT (top) and NEXT TO (bottom). Participants read one sentence (either U-E, on the top row in each trial, or E-U, on the lower row) and had to choose which of the two pictures matched that sentence. Among the prime trials, examples of U-wide primes are on the left and of E-wide primes are on the right. The right-hand picture in each pair of prime trials shows the foil, or incorrect choice. The left-hand picture shows correct prime choices (either U- or E-wide, depending on the type of prime). On target trials, the left-hand picture shows the U-wide choice and the right-hand pictures shows the E-wide choice corresponding to the example sentence.

- b. $\exists y[\text{circle}(y) \wedge \forall x[\text{star}(x) \rightarrow \text{above}(y, x)]]$
 There exists a y , such that y is a circle, and for all x , if x is a star, then y is above x

Note that the two meanings above correspond precisely to the two possible orders for applying the universal and an existential quantifier in first order logic. In other words, (2a) and (2b) differ solely in terms of which quantifier takes scope over the other. For this reason, sentences like (1) are called ‘scopally ambiguous’. Fixing the relative scope of quantifiers resolves the ambiguity by specifying the order in which the two quantifiers bind their variables. In (2a), the universal quantifier EVERY takes wide scope—for every star, that star is above a (potentially different) circle. We refer to this as a “Universal-wide” or “U-wide” interpretation. It is also the “inverse scope” reading, since the order of the quantifier at LF is the inverse of their linear order in the original sentence (1). (2b) is the interpretation where there is a circle, and all the stars are above that very same circle. This LF has the existential quantifier, A, taking wide scope, and we refer to it as an “Existential-wide” or “E-wide” interpretation. It is also the “linear scope” reading, because the relative scope of the quantifiers at LF is the same as the linear order of quantifier words in (1).

Psycholinguists have investigated how listeners derive an LF by having participants read scopally ambiguous sentences, while manipulating factors like word order and quantifier identity, and using a variety of measures to determine which reading the participants derive (Anderson, 2004; D. G. Clark & Kar, 2011; Filik, Paterson, & Liversedge, 2004; Gil, 1982; Gillen, 1991; Johnson-Laird, 1969; Kurtzman & MacDonald, 1993; Micham, Catlin, VanDerveer, & Loveland, 1980; Paterson, Filik, & Liversedge, 2008; Schlotterbeck & Bott, 2013). While this approach has helped identify factors that influence which LF gets constructed, it has not answered a question of major significance for formal semantic theory: how (and whether) the *same* LF-constructing operations act on *different* scopally ambiguous sentences —sentences which may share little in terms of their syntactic or lexical content.

Although there is a wide array of theories addressing this question, they all share one feature. They propose that when a quantifier is in object position, there are interpretive mechanisms that can assign semantic wide scope to that quantifier, thereby deriving the inverse scope reading. Recent empirical studies have looked for evidence of such an inversion operation using priming paradigms. These experiments first elicit one unambiguous LF representation (like (2a) or (2b)) for a given scopally ambiguous sentence (like (1)), and then test whether participants are more likely to arrive at a similar LF for a different ambiguous sentence. For example, if sentence (1) was the prime, sentence (3) might be the target.

(3) Every square is below a triangle.

There are two possible patterns of priming. If getting the U-wide reading of (1) makes it easier to get the U-wide reading of (3), this would suggest that what can be primed is the U-wide LF representation. To account for this priming, we would have to suppose that the representation that is being primed abstracts away from all the other differences between (1) and (3), such as the different word orders and the different nouns being quantified, but still represents the scopal relations between the universal and existential quantifiers. If, on the other hand, the U-wide reading of (1) makes the E-wide reading of (3) more available, then it would suggest that what is primed is an inversion operation that picks out the second quantifier in both sentences (again abstracting over lexical content).

This priming paradigm originated with Raffray and Pickering (2010) (henceforth, R&P). They found that forming a U-wide representation for one sentence containing EVERY and A in that order, makes it more likely that participants will form a U-wide representation for another EVERY-A sentence, even if the two contain different nouns. They also found priming from passive to active sentences, where EVERY quantified the deep subject (or agent) in both. Critical for testing an inversion operation, they found no priming effect between active sentences where the quantifiers came in opposite orders (e.g. from “A kid climbed every tree” to “Every hiker climbed a hill”). Feiman and Snedeker (2016) extended these findings, showing, with a variety of quantifiers —ALL, EACH, EVERY, and the numbers THREE, FOUR and FIVE — that U-wide representations will prime other U-wide representations when the quantifiers are the same in prime and target but not when they are different.¹ In addition, they found that this effect holds across sentences with different verbs as well as different nouns. This suggests that LF representations abstract away from both noun and verb content, but are sensitive to the differences between even quite similar quantifiers.

¹Feiman and Snedeker (2016) treat doubly-quantified sentences involving all these quantifiers (ALL, EACH, EVERY, and numerals) as instantiating a *scope* ambiguity. It is controversial, however, that the same semantic mechanism is at play in all these cases (see Chapter 7). While an inverse-scope operation directly accounts for the ambiguity arising for singular universal quantifiers (e.g. ‘each’ and ‘every’), deriving the ambiguity for sentences with plural quantifiers (e.g. ‘all’ and numerals) may require the insertion of a distributivity operator. These sentences would then be examples of *plural* ambiguities (rather than *scope* ambiguities). In any case, the conclusions drawn by Feiman and Snedeker regarding the absence of between-quantifiers priming would still hold for EACH and EVERY. Interestingly, the lack of priming between ALL and numerals could arguably suggest that priming of distributive/non-distributive readings is also quantifier specific (see Part I).

Using a different type of pictorial stimuli, involving juxtapositions of arbitrary symbols, Chemla and Bott (2015) (henceforth, C&B) tested sentences containing *EVERY* and *A*, and again found that U-wide LFs prime subsequent U-wide LFs. Unlike R&P, they fully varied the order of the two quantifier words, and found that the same scopal relation was primed in all cases: U-wide primes elicited more U-wide target readings than E-wide primes did, independently of the quantifiers' order in the two sentences. This differs from R&P, who found no priming effect of any kind when using *A-EVERY* sentences as primes and *EVERY-A* sentences as targets. We will return to these differences in the General Discussion.

At present, our goal is to follow up on C&B's positive finding, since the priming pattern they found (if it holds up) bears directly on the issue of scope inversion. Their findings suggest that LF *representations* can be primed. Had there been any priming of a scope inversion operation, U-wide readings of *A-EVERY* sentences would have increased the availability of *E-wide* readings of *EVERY-A* sentences relative to E-wide readings of *A-EVERY* sentences. The absence of this effect may suggest, as C&B in fact conclude, that the inversion operation isn't primeable, while LF representations are. But it may also bear on debates about what the inversion operation is like. While most traditional accounts posit a single quantifier-general inversion operation, which applies to any quantifier word found syntactically lower in the sentence (Chomsky, 1976; Cooper, 2013; Fodor, 1982; Jackendoff, 1972; May, 1977, 1985), other theories posit multiple different mechanisms for different quantifiers (Beghelli & Stowell, 1997; Reinhart, 1997; Steedman, 2012). If there is a different inversion operation for *EVERY* in object position than for *A*, that would explain the lack of priming between *EVERY-A* and *A-EVERY* sentences.

However, before pursuing these theoretical possibilities, there is a methodological issue with C&B's stimuli that must be addressed. As shown in the upper panel of Figure 6.1 (cf. Experiment 1), there is a strong visual similarity among all the U-wide pictures, and a different visual similarity among the E-wide pictures, deriving from the use of *ABOVE* and *BELOW* as the only two predicates across all sentences. These pictures are the same regardless of the quantifier word order, so that if there was a priming effect based on this visual similarity, it would not be affected by word order, thus mimicking the data pattern expected for the priming of LF representations. Moreover, the pictures use arbitrary symbols and the scopally ambiguous sentences describe arbitrary spatial relations between them (see Figure 6.1). This means that to match a picture to either scopal reading of the sentence, participants must check the picture's elements one by one. However, given the visual similarity of the stimuli, a simpler verification strategy is available: for U-wide primes, participants can scan the picture from side-to-side, ensuring that the central rows of elements are uniform (e.g. in the case of (1), a row of stars and a row of circles). For E-wide primes, they can scan the central column of elements up-and-down, ensuring that it is uniform up to the last element (three shapes of the same kind in a column, and a different shape on the end). Whether the predicate in the sentence is *ABOVE* or *BELOW* makes no difference for either strategy. These strategies, available in every scopally ambiguous trial, lead to the correct response, and are easier to use than a more detailed checking procedure. Critically, their use would lead to precisely the observed pattern of results—greater U-wide responding on targets following U-wide than E-wide primes, regardless of quantifier word order, or of whether the prime and target match on the predicate they use.

C&B argue that priming based on picture similarity is unlikely because it had been controlled for and ruled out in R&P's experiments. However, the stimuli used in C&B's study are quite different: they do not depict a natural scene, they use arbitrary shapes for both the subject and object of the sentence instead of distinctive agents and themes, and they relate these shapes using arbitrary spatial predicates instead of verbs with asymmetric agent-theme relationships. To be sure that there is no picture priming in C&B's study, it would be useful to control for picture similarity by using similar stimuli that do not allow for the priming of simple spatial strategies.

We report on the results from an experiment of this kind, manipulating the spatial predicate used in the primes and correspondingly changing the prime images. In Experiment 1, we replicate C&B’s priming effect when the predicates in both prime and target are ABOVE and BELOW, just as they were in the original study. This experiment differs from C&B’s only in that we have eliminated one of their sentence configurations —U-Neg sentences— which they found failed to prime anything. In Experiment 2, we change the predicates in the prime to: TO THE LEFT OF, TO THE RIGHT OF, or NEXT TO, correspondingly changing the pictures in the prime, but keeping target sentences and pictures the same as in the previous study (ABOVE and BELOW). We now find no priming effect, suggesting that C&B’s results are indeed based on picture similarity. We conclude with a discussion of what the growing priming literature has told us about the nature of LF representations and scope inversion operations so far.

6.2 Experiment 1

6.2.1 Methods

Participants

Using 80 participants, C&B found significant priming when averaging across different quantifier word orders, but not in each order separately, suggesting that more participants may be needed to find these relatively small effects. For this reason, and because we wanted to provide a strong test of the replicability of their finding, we doubled the number of participants. We recruited 176 participants on Amazon Mechanical Turk, and following C&B’s criteria, excluded 16 of them for failing more than 10% of prime trials, and one for indicating that English was not their first language. Participation was restricted to workers who had had 98% of their previous work approved, and who were logging in from an IP within the United States.

Materials

Sentences Experimental sentences were constructed according to one of two frames:

- U-E sentences: Every <shape 1> is <predicate> a <shape 2>
- E-U sentences: There is a <shape 1> is <predicate> every <shape 2>
- Filler sentences: Every <shape> is <color>

The shapes were hearts, squares, triangles, stars, diamonds or circles. The predicates were either “above” or “below”. Examples are shown in Figure 6.1 (upper panel). There were four lists of stimuli, with participants randomly assigned to a list. Each list was obtained by randomly inserting shapes into the appropriate sentence frame (with shape 1 and 2 always differing from each other). Within a list, trials were administered in random order to each participant using the Ibex platform created by Alex Drummond (<http://spellout.net/ibexfarm/>).

Images For each sentence, we constructed three types of images: a foil (F) consistent with none of the interpretations, an image consistent only with the U-wide interpretation, and an image consistent only with the E-wide interpretation. Prime trials paired a sentence with its foil image F and either a correct U-wide or E-wide image. The choice of the correct image thus forced one of the two interpretations. Target trials paired a sentence with a correct U-wide and a correct E-wide image. Participants thus chose the image corresponding to their preferred interpretation. For filler trials, just one of the two images made the sentence true.

Procedure and design

Experimental trials were presented in a prime-target pair. Primes consisted of one of two quantifier word orders (U-E or E-U), and presented a correct picture consistent with one of two interpretations (U-wide or E-wide). Targets similarly contained one of two word orders (U-E or E-U). As in C&B, the predicates in both prime and target trials were randomly and independently chosen from “above” or “below”. Thus, Experiment 1 manipulated three within-subjects factors with two levels each: Word Order in Primes (U-E or E-U); Word Order in Targets (U-E or E-U); and Prime Scope (U-wide or E-wide). A complete experimental set therefore consisted of $2 \times 2 \times 2 = 16$ prime-target pairs. We further counterbalanced the position of the correct image in prime trials and the corresponding image in target trials (left or right) to obtain 4 trials for each of the 16 experimental pairs (64 pairs). There were also 64 control trials, randomly interspersed between prime-target pairs, making the total number of trials $(64 \times 2) + 64 = 192$.

6.2.2 Results

Following C&B, we discarded responses on target trials if participants had answered the preceding prime incorrectly. This was rare (E-U primes: 1.97% U-E primes: 1.73%) Participants were on average 96% accurate on filler trials, which were not included in the analysis.

We examined whether participants chose the U-wide picture on target trials as a factor of the other variables we manipulated. Because participants' choice on each trial was binary, we analyzed the data using a logit mixed-effects model (Jaeger, 2008). The data was analyzed in the R programming language, v3.2.5 (R Core Team, 2014) using the lme4 package (Bates et al., 2014). We pursue the same analysis strategy as C&B, starting from a model with a maximal random effects structure involving all applicable random slopes (Barr et al., 2013), for each effect of interest, we report p -values from Type II Wald χ^2 tests on each factor (i.e. its significance after the inclusion of all other factors, except for higher order interactions involving that factor). We differ from C&B in reducing maximal models according to the reduction procedures recommended by Bates, Kliegl, Vasishth, and Baayen (2015).

We started with an omnibus model, including all of the variables which C&B had investigated separately: Prime Scope (U-wide vs. E-wide), Target Word Order (U-E vs. E-U), and Word Order Match (whether both prime and target sentences were either U-E or E-U or whether they differed; i.e. Matching vs. Mismatching quantifier word order), as well as their interactions. Figure 6.2 shows the results broken down by these factors. We found a significant main effect of Prime Scope ($\chi^2(1) = 79.95$, $p < 0.0001$; U-wide primes increase the probability of U-wide responses on subsequent target trials), a significant main effect of Word Order Match ($\chi^2(1) = 14.98$, $p = 0.0001$; higher rates of U-wide responses in Mismatching than Matching cases), and a significant effect of Target Word Order ($\chi^2(1) = 148.52$, $p < 0.0001$; more U-wide responses for E-U than U-E targets). There was also a significant interaction of Prime Scope and Word Order Match ($\chi^2(1) = 18.94$, $p < 0.0001$), but no interactions involving Target Word Order.

As there were no interactions with Target Word Order, we next explored the interaction between Prime Scope and Word Order Match. Dropping Target Word Order, we constructed a model including only Prime Scope (U-wide vs. E-wide) and Word Order Match (Matching vs. Mismatching). We found a main effect of Prime Scope ($\chi^2(1) = 57.55$, $p < 0.0001$), a main effect of Word Order Match ($\chi^2(1) = 10.52$, $p = 0.001$) and a significant interaction ($\chi^2(1) = 14.58$, $p = 0.0001$), reflecting a bigger priming effect on Matching targets. Simple effects revealed separately significant effects of Prime Scope for Matching ($\chi^2(1) = 85.73$, $p < 0.0001$) and Mismatching targets ($\chi^2(1) = 9.36$, $p = 0.002$). Thus, although priming was greater when word order matched between prime and target, a priming effect was present in both cases.

We wanted to further check whether the priming effect might be specific to a particular word order. Although there was no interaction between Target Word Order and Word Order Match, we also looked at the effects of Target Word Order (U-E or E-U) separately in Matching-Order

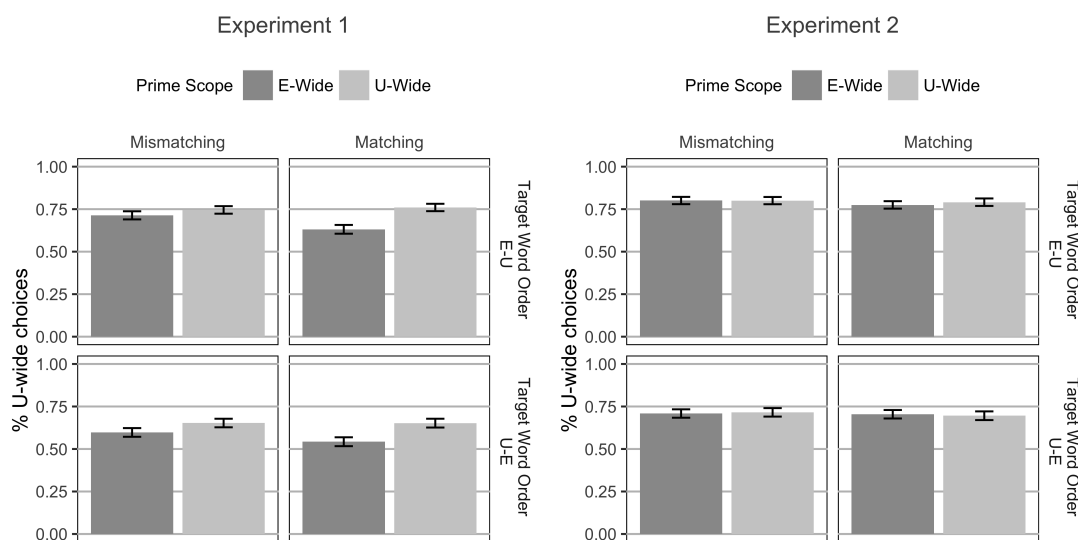


Figure 6.2: Target responses in Experiments 1 and 2. Mean percentage of U-Wide choices in targets for both prime configurations (E-Wide, U-Wide). Columns break down targets by Quantifier Word Order Match between prime and target (Mismatching, Matching). Rows show Target Quantifier Word Order (E-U, U-E)

and Mismatching-Order trials, as C&B had done. In both cases, we included Prime Scope, Target Word Order, and their interaction as predictors. In Matching-Order trials, we found a significant effect of Prime Scope ($\chi^2(1) = 77.87, p < 0.0001$; more U-wide responding after U-wide primes) and of Target Word Order ($\chi^2(1) = 57.64, p < 0.0001$), with no significant interaction. Simple effect analyses showed a significant priming effect for U-E targets ($\chi^2(1) = 28.575, p < 0.0001$) and for E-U targets ($\chi^2(1) = 56.57, p < 0.0001$). In Mismatching-Order trials, we again found a significant effect of Prime Scope ($\chi^2(1) = 13.84, p = 0.0002$) and of Target Word Order ($\chi^2(1) = 40.59, p < 0.0001$), with no significant interaction. Simple effect analyses showed a significant priming effect for U-E targets ($\chi^2(1) = 8.06, p = 0.003$) and for E-U targets ($\chi^2(1) = 4.15, p = 0.04$). Thus, priming is present regardless of the order of quantifiers either in the prime or in the target.

Finally, we also looked at priming within and between the two predicates, ABOVE and BELOW. In a model including the variables of Prime Scope and Predicate Match (Between- vs. Within-Predicate), we find a main effect of Prime Scope ($\chi^2(1) = 46.14, p < 0.0001$), no effect of Predicate Match, and no interaction ($\chi^2(1) = 0.03, p = 0.86$). Thus, the priming effect does not depend on predicate overlap. Figure 6.3 shows these results.

6.2.3 Discussion

Using identical stimuli to C&B (excluding the U-Neg sentences), we replicate their major findings. We find that U-wide prime trials increase the likelihood of U-wide target choices, both when the two sentences match in the word order of the quantifiers (U-E or E-U), and when they mismatch. This is consistent with the priming of LF representations across sentences. Unlike C&B, we also found a stronger priming effect for Matching- than Mismatching-Order prime-target pairs, consistent with the presence of an additional scope inversion operation.

However, all of these results are also consistent with a priming effect based on a form of picture similarity. That U-wide primes make U-wide targets more likely across all word orders is consistent with priming based purely on picture similarity. The fact that this priming effect is stronger in Matching-Order prime-target pairs complicates this story a bit, but does not eliminate the possibility that the priming is linked to a picture verification strategy. For example, it's possible that similarity of the sentences —by virtue of similar quantifier word orders— could affect the verification strategy used on the target trials: use of both similar sentences and

similar pictures may increase the likelihood of checking both prime and target pictures in the same way, so that the stronger priming would be due to additive effects on strategy choice.

In Experiment 2, we test these alternative explanations (scopal priming and strategy choice) by changing the predicates used in the primes. This changes not only the pictures and verification strategies on prime trials, but also their match with the targets. Note that the priming effects in Experiment 1 held whether the two sentences both contained the same predicate (either ABOVE or BELOW), suggesting that the priming effect, if not based on picture similarity, ought to be insensitive to the predicate's identity—it should generalize from LEFT to BELOW, for example, just as it did from ABOVE to BELOW.

6.3 Experiment 2

To look at the contribution of picture and verification strategy similarity, we attempted two different manipulations of the preposition in the prime sentence. In one half of the prime pictures, we used the preposition NEXT TO, which produced U-wide and E-wide pictures quite dissimilar to the target pictures (see Figure 6.1, bottom row). Whatever verification strategy participants used for these primes wouldn't readily apply to the targets, so that if such strategies drove the priming effect in Experiment 1, we would expect no priming effect here.

In the other half of the primes, we used the prepositions TO THE RIGHT/LEFT OF. As Figure 6.1 shows, this reverses the pattern of picture similarity. Where U-wide ABOVE/BELOW primes in Experiment 1 were more similar to U-wide ABOVE/BELOW targets, U-wide TO THE RIGHT/LEFT OF primes are more similar to *E-wide* ABOVE/BELOW targets. We reasoned that if similar verification strategies for similar pictures were driving the priming effect in Experiment 1 (*scan horizontally* for U-wide pictures; *scan vertically* for E-wide pictures), we would find just the reverse pattern of priming here (e.g. U-wide TO THE RIGHT/LEFT OF primes should prime E-wide readings in ABOVE/BELOW targets).

Of course, if the priming effect found by C&B, and again by us in Experiment 1, is due to the priming of LF representations, with every preferring wide scope no matter where in the sentence it is, then the specific predicates should not matter. We should find a priming effect from sentences containing TO THE RIGHT/LEFT OF or NEXT TO in the prime to sentences containing ABOVE/BELOW in the target, just as we found priming effects from ABOVE to BELOW and BELOW to ABOVE in Experiment 1.

6.3.1 Methods

Participants

Since we were interested in the possible *absence* of a priming effect in this experiment, we wanted to increase our power and thus our ability to find an effect. We again increased the number of participants, recruiting 280 participants who had not been in Experiment 1 on Amazon Mechanical Turk. Following C&B's criteria, 49 of these were excluded because they failed more than 10% of prime trials, and four were excluded because they indicated that English was not their first language.

Materials

The only change from Experiment 1 was that the prime predicates were now TO THE RIGHT/LEFT OF and NEXT TO, with the prime pictures changing correspondingly (see Figure 6.1). To avoid lengthening the experiment, we did not increase the number of trials, so that there were half as many primes with each predicate as there were with ABOVE/BELOW in Experiment 1. All other aspects of the methods were identical to Experiment 1.

6.3.2 Results and Discussion

As in Experiment 1, we first looked at an omnibus model containing Prime Scope (U-wide vs. E-wide), Target Word Order (U-E vs. E-U), and Word Order Match (Matching vs. Mismatching) and their interactions. We found significant main effects of Word Order Match ($\chi^2(1) = 8.55$, $p = 0.003$; higher U-wide response rates after mismatching than matching primes) and Target Word Order (higher U-wide response rates on E-U than U-E targets; $\chi^2(1) = 295.45$, $p < 0.0001$). Critically, unlike Experiment 1, we found no effect of Prime Scope ($\chi^2(1) = 0.41$, $p = 0.52$), and no significant interactions between any of the variables. Despite the absence of a main effect of Prime Scope, we also wanted to look at whether there might be an interaction between this variable and the Prime Predicate (RIGHT/LEFT OF vs. NEXT TO). This interaction, too, was not significant.

To check that the effect of Prime Scope in Experiment 1 was significantly different from its absence in Experiment 2, we also built a model with Experiment (1 vs. 2) as a predictor, along with Prime Scope and their interaction. As expected, we found a significant interaction of Prime Scope by Experiment ($\chi^2(1) = 36.35$, $p < 0.0001$), along with a main effect of Experiment (more U-wide responses in Experiment 2; $\chi^2(1) = 11.32$, $p = 0.0008$), and a main effect of Prime Scope (more U-wide responses after U-wide primes on average; $\chi^2(1) = 47.8$, $p < 0.0001$).

Across both experiments, we find a preference for U-wide interpretations in both EVERY-A and A-EVERY constructions. Although this may seem surprising for A-EVERY sentences, since it conflicts with the general assumption that linear scope should be easier to derive, it is consistent with many findings about the general preference of EVERY to take wide scope (see Feiman & Snedeker, 2016; Kurtzman & MacDonald, 1993). More surprising is the finding, in both experiments, that the U-wide preference is *greater* for EVERY-A than A-EVERY sentences. We know of no theoretical account that predicts this finding.

As Figure 6.3 shows, we find no difference between the predicates TO THE RIGHT/LEFT OF and NEXT TO. This is critical to interpreting the differences between Experiment 1 and Experiment 2. If participants in Experiment 2 had used the same search strategy in the target trials that they used on the prime trials (*scanning horizontally* or *scanning vertically*) it would have produced the reverse priming effect from TO THE RIGHT/LEFT OF primes to ABOVE/BELOW targets, but no priming on the NEXT TO primes. In combination with a real scopal priming effect, strategy priming could have produced a null effect for the RIGHT/LEFT primes and a priming effect for the NEXT TO primes. The absence of *any* difference suggests that there was no strategy priming in Experiment 2 at all. Given that strategy priming is presumably the cause of the effects in Experiment 1, we then have to explain why it disappears.

We see two possible explanations. First, the strategy priming may be adaptive and not merely passive: participants may only carry over strategies when there is sufficient similarity between the primes and targets to suggest that the strategy can be reused. Second, the verification strategies may be more specific and thus not apply across any of the prime-target pairs in Experiment 2. For example, they may depend not just on which direction participants had to scan, but also on whether items were lined up in a single row or column in that direction (as in E-wide cases with both types of predicates), or in two rows, in pairs lying along that direction (as in U-wide cases; see Figure 6.1). Here, LEFT and RIGHT do not match ABOVE and BELOW, since, scanning down, two rows in the former correspond to one in the latter (or vice versa, scanning side-to-side). Regardless of which of these explanations turns out to be true, the absence of priming across different predicates suggests that the presence of priming just in ABOVE/BELOW prime-target pairs depends on a high degree of picture similarity.

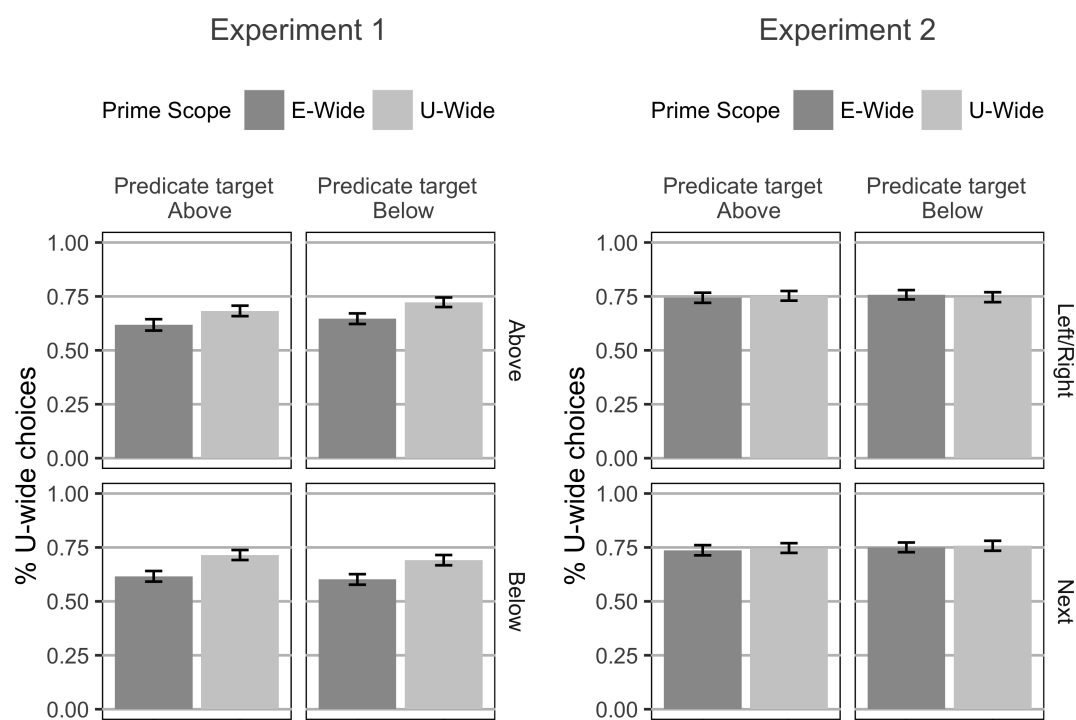


Figure 6.3: The left panel shows the effect of Prime Scope (U-wide vs. E-wide) in Experiment 1, broken down by within- vs. between-predicate. The right panel shows the effect of priming, broken down by predicate in the prime (rows: TO THE RIGHT/LEFT OF and NEXT TO) and the target (columns: ABOVE and BELOW). The Y-axis shows the percent of subsequent target trials where participants chose a U-wide picture.

6.4 Summary of Findings

The present experiments radically change the interpretation of Chemla and Bott’s (2015) findings concerning the priming of LF representations and of the operations that construct them. Experiment 1 replicates C&B, finding that U-wide primes make a U-wide interpretation of a target more likely, regardless of whether the universal quantifier comes first or second in either sentence, and regardless of whether the prime and target match in terms of its position. Experiment 2 reveals that this effect is not actually about scopal priming, but depends crucially on the similarity of pictures and strategies between the prime and target sentences. Changing the predicate in the prime in a way that forces a change in participants’ verification strategy is enough to remove the priming effect.

There is one alternative explanation that warrants further consideration. If we focus solely on the contrast between Experiments 1 and 2, we might conclude that there are scopal priming effects, but that scopal priming operations are lexically rooted and cannot abstract away from the *predicate*, just as they apparently do not abstract away from the subject *quantifier* (Feiman & Snedeker, 2016). There are two findings that argue against this possibility. First, Experiment 1 eliminates a strict lexical explanation, since it uses two predicates in both prime and target — ABOVE and BELOW — with relatively minimal changes to the corresponding pictures. Comparing within- and between-predicate priming in this case, we find no difference in the size of the priming effect. However, since ABOVE and BELOW are semantically similar (possibly even compositionally related. See H. H. Clark & Chase, 1972), this control leaves open a weaker version of the predicate hypothesis (highly similar predicates are required for priming). Stronger evidence comes from Feiman and Snedeker (2016), who used R&P’s more naturalistic set of pictures and manipulated whether the verbs in the prime and target sentences matched. They also found no difference between within- and between-verb priming. The verbs in that study

were semantically diverse (ranging from verbs of perception to verbs of motion) with no clear compositional or derivational relationships. Together, these results strongly suggest that it is picture—not predicate—similarity that is driving the priming effect in C&B’s paradigm.

There is also reason to have suspected the priming of spatially-specific verification strategies in C&B’s original data. In a third manipulation they had included—sentences involving EVERY and NOT—the corresponding pictures showed shapes that were laid out the same way for both types of prime and target pictures. The verification strategy required checking the shapes’ color rather than their spatial layout. C&B failed to find a priming effect either within EVERY-NOT sentences, or from these sentences to either U-E or E-U targets. This is exactly what we would expect if spatially-based verification strategies are the only locus of priming in this paradigm.

It is important to emphasize that the absence of semantic structural priming is specific to this paradigm. In Experiment 2, we find an absence of priming not only between different word-orders, but even within the same word-order—the case in which both Raffray and Pickering (2010) and Feiman and Snedeker (2016) had previously found significant priming effects. Unlike C&B’s study, which used different spatial configurations of arbitrary symbols, these experiments used more natural sentences (e.g. “Every kid climbed a tree”). This suggests that the type of stimuli used in scopal priming experiments matters. Complex arrangements of arbitrary symbols may lead participants to rely on verification strategies that are independent of the semantic representations of the sentence, with these strategies themselves becoming the locus of priming.

6.5 General discussion

In the introduction, we noted that scopal priming could provide evidence for a scope inversion operation. We noted that the results of C&B, taken at face value, provide evidence against either the existence or the primability of an operation of this kind. We have now demonstrated that C&B’s results are not in fact about scopal priming, but instead reflect picture verification strategies. That means we no longer have evidence that scopal priming is based only on representations rather than operations, as they had concluded. It thus makes sense to return to the psycholinguistic literature on scope inversion to see whether a full reconciliation is possible and identify the research questions that remain.

We will review the empirical literature on the existence of a general inversion operation, concluding that although there is no strong evidence of its absence, there is an absence of evidence for its existence. Furthermore, there is evidence that representation priming is quantifier-specific. Both these points suggest that inversion, too, might be sensitive to quantifier-specific properties in some way.

Psycholinguistic experiments using a priming approach provide a clear means of identifying quantifier-general operations: one can simply manipulate whether the prime and target sentences share the same quantifier words, and look at the pattern of priming effects in both conditions. In their first two experiments, (Raffray & Pickering, 2010) found a priming effect across different EVERY-A sentences (e.g. “Every kid climbed a tree”; “Every hiker climbed a hill”). Inducing U-wide readings in one sentence primed U-wide readings of the second. Moreover, the same effect was just as strong when the prime was passivized (e.g. “A tree was climbed by every kid”). However, while *passive* A-EVERY sentences primed *active* EVERY-A sentences, *active* A-EVERY sentences did not—they found no priming effect from sentences such as “A kid climbed every tree” to sentences such as “Every hiker climbed a hill”.

R&P’s data could bear on the existence of a single, quantifier-general inversion operation. Such an operation would presumably have been insensitive to the bindings between quantifiers and thematic roles, and would have given wide scope to the second quantifier regardless of its identity, resulting in U-wide readings of A-EVERY sentences priming *E-wide* readings of EVERY-A

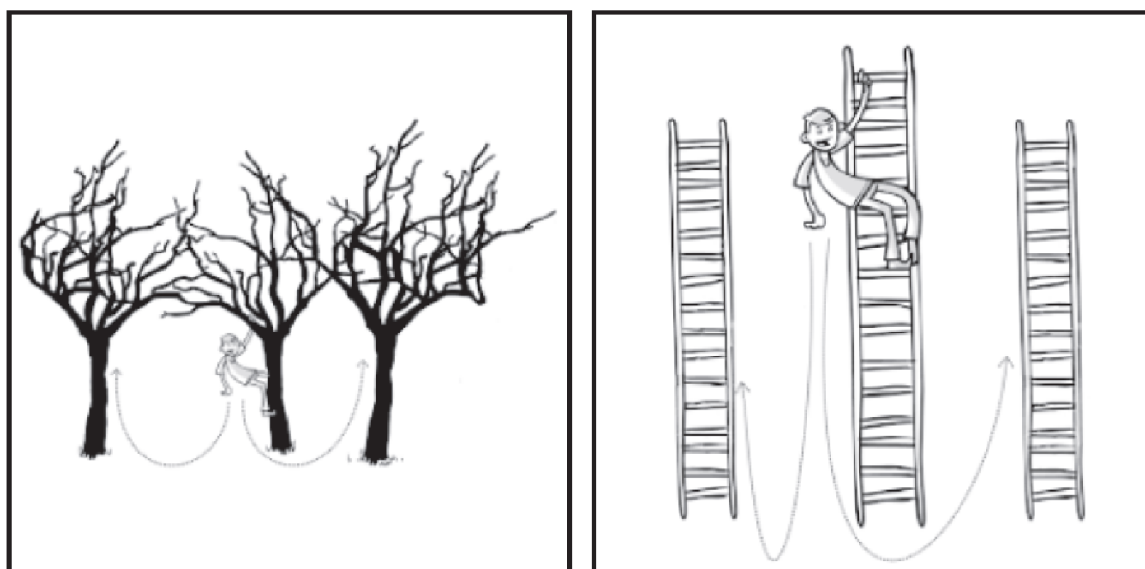


Figure 6.4: Examples of E-wide prime pictures used in [Raffray and Pickering \(2010\)](#) Experiment 3. The corresponding sentence is, “A kid climbed every tree”. Dotted lines between objects are meant to indicate a sequence of events, but may not be transparent to participants —first a kid climbed one tree (or ladder), then another, then another.

sentences. The absence of such an effect suggests the absence of a primeable quantifier-general inversion operation. However, there is a methodological issue here too, which may undermine this conclusion.

While the pictures used for prime sentences in all their other experiments are similar in form, allowing for a straightforward verification of the picture relative to one of the LFs, the prime pictures in Experiment 3 require participants to make additional assumptions that may complicate the process of interpretation. Figure 6.4 reproduces these images, showing how the E-wide prime picture in Experiment 3 requires participants to infer and assume a sequence of events —the kid first climbed one tree, then the next, then the next— before they can match this picture to an E-wide LF. Participants may have had trouble understanding that this is what the lines and arrows are meant to depict, and therefore may have failed to construct the corresponding E-wide LFs. This possibility finds support in R&P’s data, where participants’ U-wide response rates on target pictures are not only identical between E-wide and U-wide primes in Experiment 3 (74% and 76%, respectively), but also nearly identical to the response rates following U-wide primes in their Experiments 1 (77%) and 2 (74%). This suggests that it was specifically the problematic *E-wide* primes that failed to have an effect. Thus, R&P’s experiments do not rule out the possibility that an inverse reading of A-EVERY sentences can prime an inverse reading of EVERY-A, and therefore do not rule out the existence of primeable quantifier-general scope inversion.

In another priming study, [Feiman and Snedeker \(2016\)](#) looked at whether inverse scope could be primed across sentences with different quantifiers. Keeping the object quantifier as the indefinite A, they varied the quantifier in subject position —EVERY, EACH, ALL, or a number— and looked for scopal priming effects across sentences. They found significant priming effects only when the prime and target sentences shared the same quantifiers, but not when the quantifiers differed. For example, participants generally preferred an U-wide interpretation for EVERY-A sentences. U-wide primes also resulted in more U-wide interpretations for subsequent EVERY-A sentences than E-wide primes. But this priming effect was found only for EVERY-A sentences; it disappeared if the target contained any other quantifier in subject position. The same pattern of within-, but not between-quantifier priming was found for each of the other

quantifiers. This suggests that LF representations (U-wide or E-wide) can be primed, but only in a quantifier-specific way: they differentiate between quantifiers as similar as EACH and EVERY.

Like R&P, Feiman and Snedeker find no evidence of a single, quantifier-general inversion operation. Inverse readings of EACH-A sentences did not prime inverse readings of EVERY-A sentences. However, this too is not conclusive. For one, the same indefinite, A, served as the object quantifier in all of their sentences, so that any conclusion about inversion being quantifier-specific rather than quantifier-general hinges on the absence of inverse scope priming across sentences with *different* subject quantifiers. A better test of quantifier-general scope inversion would be to look at priming effects across sentences with the same subject quantifier and different object quantifiers. Additionally, studying scope inversion of the indefinite A may be irrelevant to theories that posit a quantifier-general inversion operation. The indefinite has been argued to take wide scope through a completely different mechanism that would not apply to other quantifiers in object position (Fodor & Sag, 1982; Reinhart, 1997). Thus, even if inverse readings had been primed across sentences with different subject quantifiers, the locus of priming could be this alternate mechanism.

Another less direct source of evidence about the nature of inversion comes from differences in the accessibility of linear and inverse scope readings. Five-year-old children have been shown to prefer linear scope readings across a variety of quantifier combinations in sentences containing: a universal followed by a negation, an existential followed by a negation, a negation followed by an existential, and a negation followed by a number word (Lidz & Musolino, 2002; Musolino, 1999; Viau, Lidz, & Musolino, 2010). That children assign wide scope to an existential over negation when the existential is first, but to negation over a number (essentially an existential with cardinality) when the negation is first, suggests that linear scope overrides any lexical biases children might have. While this pattern shows a bias towards linear scope interpretations at a particular point in development, it does not show that there is a single inversion operation, acting on all quantifiers in object position regardless of their identity. The linear scope bias is equally consistent with different operations for different quantifiers, which might be learned piecemeal in the course of language acquisition. It would count in support of a single operation if each child gained the ability to invert multiple different quantifiers at the exact same developmental time, but as far as we know, there is no evidence that this is the case.

What's needed now is a stronger test of an abstract inversion operation. One option is to hold the subject quantifier constant, looking for scopal priming across sentences with different object quantifiers, which are the target of any inversion operation. In doing this, there is no reason to restrict ourselves solely to scope ambiguities between universals and existentials. Scope ambiguities are also present in combinations of numerals, negation, and higher-order quantifiers like "most", "no more than four", and so on. An abstract, quantifier-general scope inversion operation, if it exists, should range across these different quantifier combinations.

6.6 Conclusion

The sum of the evidence from priming paradigms so far suggests that LF representations distinguish between different quantifiers (at least between EACH and EVERY). The prior evidence for the absence of priming of the scope-inversion operation in Chemla and Bott (2015) appears to be due to priming related to pictures and specific spatial verification strategies. Thus, the jury is still out on whether there is a general inversion operation that can be primed. It might be possible, however, to reconcile quantifier-specific differences with a common inversion operation by positing that this operation gives all quantifiers in object position higher semantic scope, but differs among them in just how high that scope is (i.e. in *where* the raised quantifier lands). Examples of such an approach have been pursued in the semantic literature by Szabolcsi (1997) and Beghelli and Stowell (1997).

7 | An experimental note on distributivity and scope*

Križ, M. and Maldonado, M. An Experimental Note on Distributivity and Scope. Manuscript.

Abstract

Experimental data are frequently used in formal semantics when intuitions about the existence of readings are shaky, and they can even help us to discover readings that go unnoticed by simple introspection (Marty, Chemla, & Spector, 2014). We report an experiment that tested the existence of distributive readings with plural noun phrases in object position (e.g. ‘A mouse is painting all the penguins’), which has been the subject of some disagreement in the literature (Steedman 2012). Our findings indicate that these readings do, indeed, exist, and we will suggest that their comparative marginality can be explained without recourse to a ban on inserting the distributivity operator in derived scope positions.

7.1 Introduction

Sentences with a definite plural noun phrase in subject position, such as (1), are truth-conditionally compatible with, roughly speaking, three types of situations:¹

- (1) The mice are painting a penguin.
 - a. Every mouse is painting one and the same penguin.
 - b. Every mouse is painting a penguin, but it’s not the same one for all of them.
 - c. The mice are collaboratively painting a penguin (jointly making the painting, etc.).

It is not so clear that the situation is analogous for sentences involving plural definites in object position. While sentence (2) is certainly compatible with scenario (2a), its status in the *co-variation* scenario in (2b) is dubious.²

- (2) A mouse is painting the penguins.
 - a. Every penguin is being painted by a single mouse.
 - b. ??Every penguin is being painted by a mouse, but it’s not the same for all of them.

The general goal of this paper is to explore experimentally the availability of different understandings for sentences such as (2), which involve a plural noun phrase in object position. Specifically, our aim is to assess whether or not these sentences can give rise to readings

We are indebted to Emmanuel Chemla for valuable comments and discussion. The research leading to this work was supported by the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.313610, and by ANR-10-IDEX-0001-02 PSL and ANR-10-LABX-0087 IEC.

¹We are simplifying here by ignoring situations that are, as it were, in between the latter two, such as several teams of mice each collaboratively painting one and the same penguin.

²There is not analogue of situation (1c) here because every collective being-painted is also an individual being-painted, so that there is no genuinely collective reading with respect to the object position of *paint*.

compatible with co-variation scenarios. Intuitions about such interpretations seem to be shaky, but experimental data can help us to discover readings that go unnoticed by simple introspection (Marty et al., 2014). Before introducing more details about the experimental set-up, we need to provide some formal semantics background on the ambiguities that arise with plural expressions.

7.1.1 The distributive/non-distributive contrast.

Most semantic approaches to plurality assume that sentences such as (1) can have two alternative readings. A first *non-distributive* interpretation is obtained *by default*, as soon as the predicate is applied to the plurality (see (3)).

- (3) $\exists y.\text{penguin}(y) \wedge \text{is-painting}(\iota x.\text{mice}'(x), y)$
 There is a penguin, call it y , and a plurality made up of mice, call it X , such that X are painting y .

This reading is unquestionably true in the collective situation (1c). Whether it is also true in (1a) depends on what exactly one takes the predicate meaning to be. It is commonly assumed (Scha, 1981, a.m.o.) that if a predicate or relation is true of individuals, it is also true of the plurality made up of those individuals. Under this view, the reading in (3) is true both in (1a) and (1c), since it does not matter whether the individuals that make up the plurality are collaborating on the painting or are doing it independently. Crucially, this reading is not true in (1b): since the existential quantifier over penguins is the only quantifier in this reading, it has to be one and the same penguin — it contains only one quantificational operator, the existential, hence the penguin cannot vary.

To account for the reading that is compatible with the co-variation scenario in (1b), a second, *distributive* reading is assumed to come about through the insertion of an unpronounced distributivity operator D at the level of the verb phrase. This operator, which can be thought of as a silent version of the English word *each*, takes as its argument a predicate P and returns a new predicate that is true of a plurality if and only if P is true of all the atomic individuals that make up that plurality:

- (4) The mice D are painting a penguin.
 $\forall x.x \preceq_{AT} \iota z.\text{mice}(z) \rightarrow \exists y.\text{penguin}(y) \wedge \text{is-painting}(x, y)$
 For each atomic member x of the maximal plurality of mice, there is a penguin y such that x is painting y .

Since the existential quantifier *a penguin* is in the scope of the distributivity operator, which implicitly universally quantifies over atomic individuals, the penguin is allowed to vary by mouse (although it does not have to).

7.1.2 Plural quantifiers.

The observation from (1) holds not only with definite plurals, but with all plural DPs, including those headed by numerals and *all*. The sentences in (5), for example, are compatible with the same kinds of situations as (1); namely, with both *covariation* (where penguins vary by mouse) and *non-covariation* (where there is only one penguin) scenarios.

- (5) a. Two mice are painting a penguin.
 b. All the mice are painting a penguin.

It is natural to assume that the sentences in (5) are ambiguous in the same way as (1): as long as both *all* and numerals quantify over pluralities, (5a) and (5b) can receive their non-distributive reading by default, and their distributive reading — compatible with covariation

scenarios — through insertion of the *D* operator. For instance, in (5b), the indefinite *two mice* asserts the existence of a plurality of two mice, of which the predicate *are painting a penguin* can be true either non-distributively (6a) or distributively (6b), depending on whether or not the *D* operator has been inserted:³

- (6) a. Two mice are painting a penguin.
 $\exists x.\text{mice}(x) \wedge |x| = 2 \wedge \exists y.\text{penguin}(y) \wedge \text{painting}'(x, y)$
 There is a plurality made of two mice *X* and a penguin *y* such that *X* are painting *y*.
 b. Two mice *D* are painting a penguin.
 $\exists x.\text{mice}(x) \wedge |x| = 2 \wedge D(\lambda z.\exists y.\text{penguin}(y) \wedge \text{painting}'(z, y))(x)$
 There is a plurality made of two mice *X* such that, for each atomic part *x* of *X*, there is a penguin *y* such that *x* is painting *y*.

This distributive/non-distributive distinction is reminiscent of, but not identical to, the *scope ambiguities* that arise with double-quantified sentences such as (7).

- (7) Every mouse is painting a penguin.
 a. $\forall x.\text{mice}(x) \rightarrow \exists y.\text{penguin}(y) \wedge \text{painting}(x, y)$
 For every mouse *x*, there is a (potentially different) penguin *x* such that *y* is painting *x*.
 b. $\exists y.\text{penguin}'(y) \wedge \forall x.\text{mice}'(x) \rightarrow \text{painting}(x, y)$
 There is a single penguin *x* such that for every mouse *x*, *y* is painting *x*.

Sentence (7) has two different readings, (7a) and (7b), depending on the relative scope of the quantifiers *every* and *a* (Fox, 2000; May, 1985, a.m.o). In (7a), the universal quantifier *every* takes wide scope above the existential, matching the linear order of the sentence (*surface-scope reading*). Instead, in (7b), the scope relation is not isomorphic to linear order: the existential takes wide-scope above the universal (*inverse-scope reading*). Surface-scope interpretations are considered to be basic, whereas inverse-scope readings are often assumed to be derived by applying a *scope-inverting* operation, which reverses the relative scope of the quantifiers at the interpretation stage.⁴

Like the quantified-sentences in (5), scope-ambiguous sentences like (7) are also compatible with *covariation* and *non-covariation* scenarios. One might thus come to conclude that this is really the same phenomenon, but there are good reasons to believe that this is not the case. One of the main arguments comes from differences in the semantics of plural and singular quantifiers (Champollion, 2010; Dowty et al., 1987; Winter, 2001). Unlike morphologically singular quantifiers (e.g. *every*), plural quantifiers are compatible with necessarily collective verb phrases⁵.

- (8) a. All the mice are painting a penguin together.
 b. #Every mouse is painting a penguin together.
 (9) a. All the mice are gathering in the hallway.
 b. #Every mouse is gathering in the hallway.

³In the case of the universal *all*, we should note that, as far as truth conditions are concerned, *all the mice* has actually the same meaning as *the mice*. The semantic effect of *all* is of a different sort, such as adding some additional constraints on the predicates that noun phrase may combine with (Champollion, 2010), eliminating the possibility of non-maximal interpretations of the definite plural (Brisson, 1998; Križ, 2015), or removing logical trivalence effects (Križ, 2015).

⁴The need for a scope-inverting operation is uncontroversial, but there are different possible implementations: this operation has been viewed as syntactic movement (Fox, 2000; May, 1985; Montague, 1972), semantic type-shifting, or a more complex mapping at the syntax-semantics interface (Beghelli & Stowell, 1997; Reinhart, 1997; Steedman, 2012).

⁵Indeed, the sentences in (5) can also have *collective* understandings, which do not exist for scope-ambiguous sentences with singular quantifiers like (7).

This suggests that singular quantifiers are *inherently* distributive, quantifying directly over atomic individuals, and therefore incompatible with collective predicates. Plural quantifiers, on the other hand, quantify over pluralities. As a result, the *D*-operator must be inserted to derive distributive interpretations compatible with *covariation* scenarios. For the sake of brevity, we will not further explore the details of this argument. The reader is referred to Winter (2001) and Champollion (2010) to a deeper discussion of these differences.

‘Object distributivity’: inverse-scope distributive readings. The existence of a mechanism such as the distributivity operator directly accounts for ‘subject-distributivity’: the predicate is distributed over each member of the plural subject, allowing objects to covary (e.g. (1) and (5)). It has been a matter of some debate whether the same mechanism can derive what we could call ‘object-distributivity’, namely, a reading that allows the *subject* to covary with each member of the plural *object* (for an overview, see Steedman, 2012: §3.1–3.3). Such reading would make the sentence (2) (repeated in (10)) compatible with a *covariation* scenario, where there is a different mouse painting each penguin (see scenario in (2b)).

The derivation of ‘object-distributivity’ would require applying the scope-inverting operation before inserting the *D* operator: the plural noun phrase should raise and the *D* operator should be inserted below its derived position, as it is schematically shown in (10a). The resulting truth-conditions are given in (10b).

- (10) A mouse is painting the/all the/two penguins.
 a. ??the/all the/two penguins *D* λx a mouse are painting x .
 b. ??‘Every penguin is being painted by (potentially varying) mice.’

Such *inverse-scope distributive readings* are difficult to access introspectively. This is quite different from the scope ambiguity of the sentence that is obtained by replacing the definite plural with a genuine universal quantifier: in (10), it is quite apparent that the sentence has a reading where *every penguin* takes scope of *a mouse* and therefore allows covariation of mice with penguins.

- (11) A mouse is painting every penguin.
 For every penguin y , there is a (potentially different) mouse x such that x is painting y .

The apparent introspective unavailability of *inverse-distributive readings* could have two different sources. It is possible that these readings are simply not generated by the grammar, for example, because the *D*-operator cannot be inserted in derived scope positions (Steedman, 2012). Alternatively, the apparent unavailability might be the reflex of processing cost or dispreference, resulting from the additive effect of applying two different interpretative operations during parsing (i.e. scope-inversion and insertion of *D*).

7.1.3 Previous psycholinguistic work.

The question of whether alternative readings of ambiguous sentences differ in terms of preference or cognitive cost has been a focus of attention in psycholinguistic research for some time. For distributive/non-distributive ambiguities, several studies have shown that distributive readings of sentences such as (1) tend to be both dispreferred and costly in comparison to non-distributive ones (Brasoveanu & Dotlačil, 2015; Brooks & Braine, 1996; Dotlačil, 2010; Kaup et al., 2002; Musolino, 2009; Syrett & Musolino, 2013; Ussery, 1998). Recent priming studies have also revealed an asymmetry between distributive and non-distributive readings (Maldonado, Chemla, & Spector, 2017), suggesting that these differences in parsing have a semantic source. These findings are well accounted for by the picture we described above, in which non-distributive readings are primitives, while distributive readings are derived through an additional operation (in this case, insertion of *D*).

The general preference for non-distributive readings was found for sentences involving a range of different plural expressions. However, it has been suggested that the degree to which distributive interpretations are dispreferred or costly might also be partially determined by lexical properties of the quantifier, among other factors (Dotlacil, 2010; see also Part I. This suggestion, however, mostly comes from a *between-studies* comparison, and it has not been directly tested.

Experimental literature on scope ambiguities has also suggested that scope preferences have a lexical component, such that quantifiers differ on the scope they are likely to take (wide vs. narrow; Feiman & Snedeker, 2016; Ioup, 1995; Turnstall, 1998). The universal quantifier *each*, for instance, seems to have a stronger preference for taking wide scope than the quantifier *every*⁶.

These quantifier-specific scope preferences seem to be partially independent of whether the relative scope assignment is isomorphic to linear order. For example, many studies have found that *every* displays a general preference to take wide scope over *a* for both possible linear orders (O. Bott & Radó, 2009; Chemla & Bott, 2015; Raffray & Pickering, 2010; and see results and discussion in Chapter 6. The scope-inverting operation, however, might still have an effect on parsing, modulating quantifier-specific preferences (Anderson, 2004; Lidz & Musolino, 2002; Viau et al., 2010; see Pyllkänen & McElree, 2006 and Feiman & Snedeker, 2016 for discussion).

The status of *inverse-distributive readings* remains mostly unexplored in the literature, with the exception of Gil (1982). In an early study, Gil tested the availability of different readings of sentences involving numeric expressions (e.g. ‘Two boys saw three girls’), and suggested that inverse-distributive readings are mostly unavailable for these sentences.

This paper aims at testing whether *inverse distributive readings* (i.e. ‘object-distributivity’) for sentences involving different plural expressions can be accessed by speakers. If these readings are marginally available, it would suggest a specific interaction between two distinct phenomena, distributivity and scope-inversion, which remains mostly unexplored experimentally. As a secondary goal, we aim to assess to which extent the availability of these readings might depend on the specific quantifier. The degree to which these readings are accepted might vary across plural quantifiers, depending on quantifier-specific scope and distributivity preferences. The outcome for different plural expressions might therefore shed light on their lexical properties.

7.2 Methods

7.2.1 Participants

We recruited 163 subjects via Amazon’s Mechanical Turk, three of whom were excluded from the analysis for not being a native speaker of English. IPs were restricted to the United States and participants were paid 1.5 USD for their participation. The experiment itself made use of the Ibex platform (<http://spellout.net/ibexfarm/>).

7.2.2 Task and Design

Participants were instructed to perform a Truth-Value Judgment Task (TVJT). Each item consisted of a picture and a sentence (an example is given in Figure 7.1), and subjects had to judge the sentence as either *true* or *false* with respect to the picture.

The experimental design involved three fully crossed factors, FRAME, DETERMINER, and TRUTH, for a total of $2 \times 3 \times 3$ conditions. The sentences could follow one of two FRAMES

⁶Some studies that compare quantifier-specific scope preferences have also included sentences such as (5a) and (5b) (Feiman & Snedeker, 2016), treating them as scope ambiguous sentences. As expected, the preferred reading for these cases is the non-distributive one, which in these studies was taken to be a preference for inverse-scope interpretations.

A mouse is painting the penguins.



True

False

Figure 7.1: Example item from the INVERSE/THE/BOTH condition.

depending on whether the DP of interest was in subject (SURFACE frame) or object (INVERSE frame) position. The other argument position was always filled by a singular indefinite.

- (12) a. DET <animals> ... an <animal>. (SURFACE)
 b. An <animal> ... DET <animals> (INVERSE)

The DETERMINER could be either *all the*, plain *the*, or *two*. All verbs were transitive eventive predicates, and each predicate came associated with a subject and an object animal species (for example, when the predicate was *paint*, it was always mice doing the painting and penguins being painted).

All of these sentences have two potential readings, a distributive reading and a non-distributive one. In the SURFACE frame, the distributive reading is a regular instance of distributivity. The distributive reading in the INVERSE frame is the one whose existence is in question (e.g. *inverse distributive readings*). (13) shows the example of an INVERSE/ALL sentence with its two potential readings.

- (13) A mouse is painting all the penguins.
 a. Every penguin is being painted by a (possibly different) mouse. (DIST)
 b. The same one mouse is painting all the penguins. (BOTH)

The factor TRUTH determined which of these readings was true given the picture. In the NEITHER conditions, neither reading was true; in the DIST conditions, only the distributive reading was true; and in the BOTH conditions, both readings were true. Note that the pictures in the latter conditions were non-distributive, but also non-collective (no collaborative painting, see (1a)). The distributive reading was therefore true because “one and the same mouse” is just a special case of “a (possibly different) mouse”.

There were three types of pictures, and the factors FRAME and TRUTH together determined which one was instantiated: A_1VB_n , A_nVB_1 , and A_nVB_n (Table 7.1). Here, A is to be understood as the restrictor noun of the subject noun phrase, B as the restrictor noun of the object noun phrase, and V as the transitive predicate. A picture of type A_1VB_n had one A standing in relation V to n B s, and an additional A that did not stand in the relation to anyone. Pictures of type A_nVB_n had n A s and n B s in a one-to-one V relation. Pictures of type A_nVB_1 had n A s all standing in relation V to the same one B , and an additional B that did not stand in the relation to anyone.⁷ When the determiner was ALL, then n was set to 3, otherwise it was 2. The reason

⁷The additional bystander A in the A_1VB_n pictures served the purpose of satisfying the plurality presupposition

for this was to fulfil the *more than two* antipresupposition for all' (Heim, 1991; Percus, 2006). Examples of the pictures for a $n = 2$ are shown in Figure 7.2. Table 7.1 shows how FRAME and TRUTH determined a picture type.

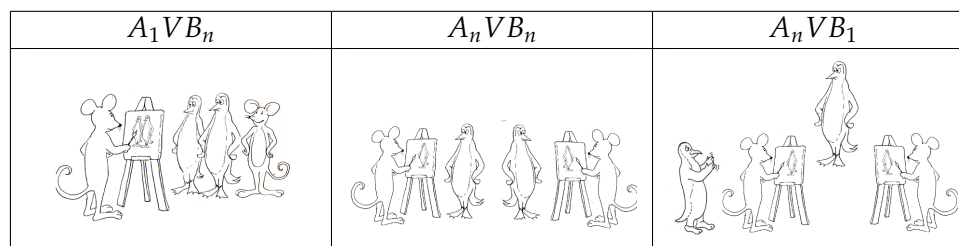


Figure 7.2: Picture types with example for $n = 2$.

FRAME	SENTENCE	TRUTH		
		NEITHER	DIST	BOTH
SURFACE	All the As V a B .	A_1VB_n	A_nVB_n	A_nVB_1
	The As V a B .			
	Two As V a B .			
INVERSE	An A V all the Bs .	A_nVB_1	A_nVB_n	A_1VB_n
	An A V the Bs .			
	An A V two Bs .			

Table 7.1: Determination of picture type by FRAME and TRUTH.

Each subject saw five instances of each condition (with different verbs and animal species) for a total of 60 items per participant. Materials, data and analyses are available online in [this link](#).

7.3 Results

Percentage of ‘true’ responses for all values of DETERMINER, FRAME, and TRUTH are shown in Figure 7.3. Our analyses were performed under two basic considerations. First, to the extent a distributive reading does not exist, *dist* conditions (in blue) should behave like *neither* conditions (in orange) within a frame. Second, if inverse-distributive readings are less available than basic distributive readings, the *dist* condition should be lower in the inverse than in the surface frame. These comparisons form the basis of our analyses.

Now, given that we are comparing three different noun phrases, we wish to ask the question of “Which, if any, of *all*, *the* and *two* behave the same?”. A null-hypothesis testing framework does not allow one to directly address a question of this form. For this reason, we decided to perform Bayesian analyses of the data. We used STAN/rstanarm to fit a variety of mixed-effects logit models and compared these models with or without quantifier-specific factors via leave-one-out cross-validation (Vehtari, Gelman, & Gabry, 2016). For the sake of brevity, the details of our analyses are provided in the Supplementary Results section (Section 7.6). The general results of our analyses are the following:

1. Distributive readings are clearly more available with *all* than with the other quantifiers ($\Delta_{elpd} = -83.6$ in *surface* and $\Delta_{elpd} = -11$ in *inverse* frame).
2. There is negligible evidence for a difference between *the* and *two* in the availability of distributive readings ($\Delta_{elpd} = -0.4$ in *surface* and $\Delta_{elpd} = -3.0$ in *inverse* frame, with the effects going in opposite directions).

of (all) the As in the SURFACE conditions, and analogously the additional B in the A_nVB_1 pictures.

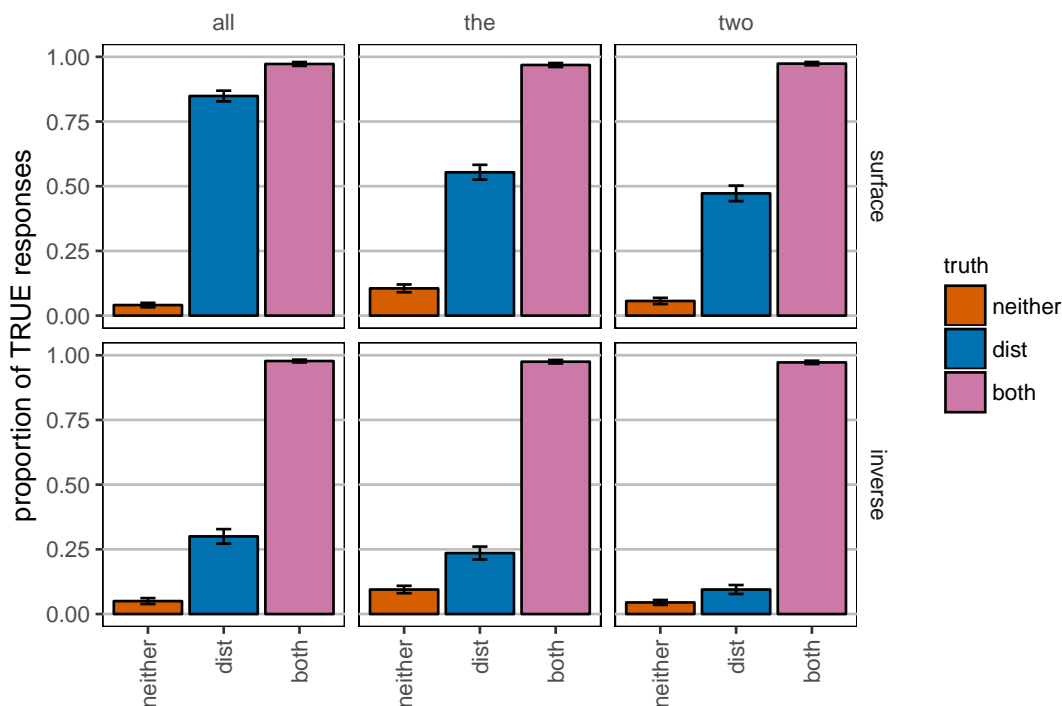


Figure 7.3: Mean proportions of *true* responses in all conditions. Error bars correspond to standard error of the mean per participant.

3. Evidence for inverse scope distributive readings with all quantifiers ($\Delta_{elpd} = -165.7$ for *all*, $-37.$ for *the*, -8.1 for *two*).
4. Assuming that the availability of inverse scope distributive readings is modulated by (i) the quantifier’s inherent preference for distributive readings (as in 1. above) and (ii) the quantifier’s scope preferences, we find models ranked as follows with respect to quantifiers’ propensity for inverse scope: *the* > *two* > *all* is better than *the* > *all* = *two* ($\Delta_{elpd} = -4.9$, $se(\Delta_{elpd}) = 3.5$), which is better than *the* = *two* > *all* ($\Delta_{elpd} = -3.0$, $se(\Delta_{elpd}) = 6.9$). The strength of the evidence here is not overwhelming, and so the picture is overall somewhat muddy, but it is worth pointing out the following: among these three determiners, *all* actually has the lowest propensity for inverse scope, despite the fact that inverse-scope distributive readings are most easily available with *all*.

Note also that it is clear that idiosyncratic scope preferences do exist. A model with only an across-the-board factor for inverse scope that was not quantifier-specific was markedly worse than the next best model ($\Delta_{elpd} = 22.2$, $se(\Delta_{elpd}) = 6.9$).

7.4 Discussion

We have provided counter-evidence to the common assumption that inverse scope distributive readings with plural quantifiers do not exist. These readings are shown to be marginally available for *all*, definite descriptions and numerals, suggesting that it is possible to insert the distributivity operator in derived-scope positions.

The marginal status of these readings can be understood as resulting from the interaction of distributivity and scope-inversion. That is, low acceptability rates for inverse distributive readings could be indicative of dispreference or high processing cost (for this particular reading), and these would in turn be caused by the combination of two different interpretative

mechanisms. The degree to which inverse distributive readings are available, however, also seems to be modulated by quantifier-specific preferences for both distributivity and scope.

First, we found evidence for lexically conditioned differences in the availability of distributive readings. While non-distributive readings are universally accepted⁸, distributive interpretations are in general more readily available with *all* than with definites and numerals.

These findings pattern with previous results in the literature, where the relative preference for distributive readings was found to be higher for *all* than for numerals (Feiman & Snedeker, 2016). On the assumption that distributive readings are always derived by applying the *D* operator, the difference between determiners is presumably grounded in a frequency effect: the probability that the silent *D* operator is there might be higher after *all* than after *the* and *two*, making the parsing with *D* more likely in the former case.

Second, we also found some evidence for a difference in the propensity of plural noun phrases to take inverse scope. Noun phrases headed by *the* are the most likely ones to take inverse-scope. On the other end of the hierarchy, noun phrases with *all* have the lowest propensity for inverse scope. Interestingly, quantifier-specific scope preferences of this sort have been argued to exist for other quantifiers (Feiman & Snedeker, 2016; Ioup, 1995; Turnstall, 1998).

Overall, *all* is still the determiner with which inverse-scope distributive readings are most available, but that is purely because its preference for distributive readings is much stronger than with other determiners, and it is just attenuated, but not nullified, by the greater dispreference for inverse scope. In other words, *all* has a higher preference for distributivity than a dispreference for inverse scope.

Taken together, our findings suggest not only that scope and distributivity do interact — giving rise to inverse-distributive readings —, but also that the rate at which these phenomena occur is at least partly dependent on specific biases of individual quantifiers.⁹ The fact that quantifiers' preferences play a role in the availability of distributive and inverse-scope interpretations has practical implications for both experimental and intuition-based approaches to these phenomena, as general claims about distributivity and scope should be tested for a range of different determiners.

Before concluding, let us note an alternative explanation of our findings, whereby the asymmetry in the availability of inverse distributive readings between *all*, on the one hand, and *the* and *two*, on the other, has a different source than a lexical preference. Naturally, this would be independent of the fact that inverse distributive readings do exist for the three plural noun phrases —i.e. they are also marginally available for definites and numerals.

One could imagine, for instance, that *all* obtains its distributive readings in a different way from definite plurals and numerals. The most natural form this view could take is to assign to *all* its regular universal quantifier meaning and assume that there is a collectivising operator that can be applied to it to obtain collective readings (see (8) and (9)).¹⁰ *All* is then, in a sense, the mirror image of definite plurals: There is no operator in distributive readings, but there is one in non-distributive readings — exactly the opposite of what is assumed for definite plurals. Distributive interpretations would then be obtained by default for *all*, explaining

⁸Let us note as a secondary point that the fact that non-distributive readings are fully available supports the assumption of closure under fusion for lexical predicates, which predicts that the non-distributive reading is made true equally by collective and separate individual actions ((1a) and (1c) from the beginning of this article).

⁹Some authors have suggested that differences in quantifiers' scope-taking behaviour might not reflect a simple frequency effect, but a deeper difference in the nature of the scope-inversion operation (Feiman & Snedeker, 2016). Instead of positing a unique scope-inversion operation for all quantifiers, these approaches have suggested that quantifiers might differ from each other in their scopal mechanism (in the lines of Beghelli and Stowell (1997) and Steedman (2012)). At the moment, however, there is not enough evidence to tease these options apart. Given that all the noun phrases tested in this paper can take inverse scope, we maintain the general view of a single mechanism for all quantifiers.

¹⁰Such a picture is argued for by Winter (2001) for independent reasons — his system also predicts inverse scope distributive readings for definite plurals and numerals, but does so via a different mechanism that is independent of the meaning of *all*.

why these readings are more available than for the other determiners (in both surface and inverse frames).¹¹ Potential counterarguments to this view come from the fact that *all* does not consistently contrast with *two* and *the*, as observed for, for instance, scope preferences.

7.5 Conclusions

This paper tested the existence of *inverse distributive readings* for sentences involving plural noun phrases in object position. By using a range of plural quantifiers, we additionally explored how lexical preferences might modulate the propensity to take distributive and inverse-scope interpretations. Our findings make two main contributions. First, we found that, despite being quite marginal, inverse-distributive readings exist for *all*, definite descriptions and numerals. Moreover, the availability of these readings is shown to be partly conditioned by quantifier preferences regarding both distributivity and scope.

7.6 Supplement on statistical analysis

Since we are comparing three types of noun phrases, the questions we wish to ask all have the fundamental form “Which, if any, of these three are the same?”. A null-hypothesis testing framework does not allow one to straightforwardly address a question of this form. For this reason, among others, we opted to conduct a Bayesian analysis. Mixed-effects logit models were fitted with STAN using the `rstanarm` library. The prior distribution for all parameters was Student’s *t* distribution with mean 0 and 5 degrees of freedom.¹² 10,000 samples of the likelihood of each data point were drawn after 10,000 burn-in iterations from four chains, for a total of 40,000 samples. Models were evaluated by leave-one-out cross-validation, approximated by Pareto-smoothed importance sampling with the `loo` package (Vehtari et al. 2016) based on the resulting 40,000 posterior samples.

Central to our analysis are the following two considerations.

- To the extent that distributive readings do not exist, the DIST conditions (in a given frame) should behave like the NEITHER conditions.
- To the extent that acceptance in the DIST/INV condition is lower than in the DIST

Our analysis was built on the following two considerations. First, to the extent that inverse-scope distributive readings do not exist, the DIST/INV conditions should behave like the NEITHER conditions for the same determiner. Second,

In order to avoid having to fit an excessive number of models, our analysis proceeded in steps corresponding to a number of questions that are interesting to ask about our data.

7.6.1 Analysis 1: Baseline

For each determiner, the *neither* conditions serve as a baseline, in that to the extent that distributive readings don’t exist with that determiner, the DIST conditions should look just like the NEITHER conditions.

To reduce the number of models to be computed and compared later on, we first restricted our data to the NEITHER conditions (from both frames, since there is no reason to expect a difference between the frames here) and performed inference on whether the baseline is the same for all three determiners.

¹¹The still high rate of non-distributive readings for *all* in the surface scope position could be understood by positing that there are multiple ways to arrive at such non-distributive interpretation.

¹²We spot-checked by fitting a few models with slightly different prior distributions and found our results did not change.

In particular, we fitted mixed effects logit models of the following general form:

$$Y_{isd} \sim \text{bernoulli}(\text{logit}^{-1}(\pi_{isd}))$$

with: $\pi_{isd} = \alpha_d + u_{1s}$

where s was the subject and d the determiner. Five models were fitted which differed in which, if any, of the α -parameters were the same. Results of the leave-one-out cross-validation are shown in Table 7.2. The model where *all* and *two* are the same, but *the* is different, is roughly equivalent to the model where all three are different ($\Delta_{\text{elpd}} = -0.6$, $se(\Delta_{\text{elpd}}) = 0.7$), and there is substantial evidence in its favour compared to the next best model, where *the* and *two* are equal and *all* is different ($\Delta_{\text{elpd}} = -11.7$, $se(\Delta_{\text{elpd}}) = 5.3$).

Parameters	elpd	se(elpd)
$\alpha_{\text{the}}, \alpha_{\text{all}} = \alpha_{\text{two}}$	-1072.9	41.2
$\alpha_{\text{the}}, \alpha_{\text{all}}, \alpha_{\text{two}}$	-1073.5	41.2
$\alpha_{\text{the}} = \alpha_{\text{two}}, \alpha_{\text{all}}$	-1084.6	41.5
$\alpha_{\text{the}} = \alpha_{\text{all}}, \alpha_{\text{two}}$	-1088.6	41.7
$\alpha_{\text{the}} = \alpha_{\text{all}} = \alpha_{\text{two}}$	-1090.5	41.9

Table 7.2: Estimated log predictive likelihood for models in Analysis 1.

This is in accordance with theoretical expectations: sentences with *all* and *two* are simply false in the NEITHER conditions, whereas those with *the* suffer a homogeneity violation rather than being simply false. The latter would plausibly translate to a slightly higher rate of acceptance (Schwarz 2013), but there is little reason a priori to expect a difference between *all* and *two*.

To simplify matters, we therefore initially adopted the assumption that $\alpha_{\text{all}} = \alpha_{\text{two}}$ as the basis of our further analyses. A later batch of models where this assumption was dropped confirmed the results described here.

7.6.2 Analysis 2: Distributivity preferences

This analysis is concerned with the availability of distributive readings, and the extent to which this differs between determiners.

Analysis 2a: Surface frame

Parameters	elpd	se(elpd)
$\beta_{\text{all}}, \beta_{\text{the}} = \beta_{\text{two}}$	-1698.2	41.4
$\beta_{\text{all}}, \beta_{\text{the}}, \beta_{\text{two}}$	-1698.6	41.3
$\beta_{\text{all}} = \beta_{\text{the}}, \beta_{\text{two}}$	-1781.8	42.3
$\beta_{\text{all}} = \beta_{\text{two}}, \beta_{\text{the}}$	-1884.5	39.0
$\beta_{\text{all}} = \beta_{\text{the}} = \beta_{\text{two}}$	-1907.9	39.5

Table 7.3: Estimated log predictive likelihood for models in Analysis 2a.

We first we restricted the data set to items from the SURFACE frame in the NEITHER or DIST conditions (thus excluding BOTH conditions, which were supposed to be, and were indeed, at ceiling) and fitted mixed effects logit models of the following general form:

$$Y_{isd} \sim \text{bernoulli}(\text{logit}^{-1}(\pi_{isd}))$$

with: $\pi_{isd} = \alpha_d + u_{1s} + (\beta_d + u_{2s}) \cdot \text{DIST}_i$

where s was the subject and d the determiner. DIST was 0 or 1 depending on whether the item was from the NEITHER or the DIST condition. As per Analysis 1, α_{all} and α_{two} were constrained to be identical.

The β -parameters are a measure of the preference of the distributive reading relative to the non-distributive reading. We fitted five models which varied in which, if any, of the three β -parameters were set to be identical. The results we obtained lead to two conclusions:

- Distributive readings are beyond doubt more preferred with *all* than with the other two types of DPs ($\Delta_{elpd} = -83.6$, $se(\Delta_{elpd}) = 12.8$).
- There is no evidence for a difference in the availability of distributive readings between the definite plural and the numeral ($\Delta_{elpd} = -0.4$, $se(\Delta_{elpd}) = 1.4$).

Analysis 2b: Inverse frame

Parameters	elpd	se(elpd)
$\beta_{all}, \beta_{the}, \beta_{two}$	-1421.2	41.2
$\beta_{all}, \beta_{the} = \beta_{two}$	-1424.3	41.2
$\beta_{all} = \beta_{the}, \beta_{two}$	-1435.3	41.4
$\beta_{all} = \beta_{two}, \beta_{the}$	-1494.9	41.2
$\beta_{all} = \beta_{the} = \beta_{two}$	-1496.3	41.2

Table 7.4: Estimated log predictive likelihood for models in Analysis 2b.

This analysis mirrors that of the previous section, but applied to the INVERSE frame. The results are shown in Table 7.4.

- There is dubious evidence for a difference in the availability of inverse-scope distributive readings for all three quantifiers ($\Delta_{elpd} = -3.0$, $se(\Delta_{elpd}) = 3.1$)
- If only one quantifier is different from the other two, then it is probably *all*, as in the surface conditions ($\Delta_{elpd} = -11.0$, $se(\Delta_{elpd}) = 8.1$).

Analysis 2c: Inverse-distributive readings

Parameters	elpd	se(elpd)
$\beta_{all}, \beta_{the}, \beta_{two}$	-1476.3	41.5
$\beta_{all}, \beta_{the} = \beta_{two}$	-1479.0	41.5
β_{all}, β_{the}	-1487.1	41.3
β_{all}, β_{two}	-1516.9	41.1
β_{all}	-1527.9	40.9
β_{the}, β_{two}	-1644.7	41.2

Table 7.5: Estimated log predictive likelihood for models in Analysis 2c. β -parameters set to 0 are omitted.

The analysis in the previous section targets differences in the availability of inverse-scope distributive readings between determiners, assuming the possibility of such readings. We therefore still need to establish whether, and for which determiners, such readings do, in fact, exist.

We can take the best model or models from the previous section and selectively set the β -parameters to 0 for certain determiners. To the extent that those models do worse than

the original ones, we have evidence for the genuine existence of inverse-scope distributive readings.¹³

Table 7.5 shows the estimated log predictive likelihoods for various models. We find strong evidence that inverse-scope distributive readings do, indeed, exist for all three determiners, with the evidence being weakest, but still not negligible, for the case of *two* ($\Delta_{\text{elpd}} = -8.1$, $se(\Delta_{\text{elpd}}) = 6.9$).

7.6.3 Analysis 3: Scope-taking preferences

Parameters	elpd	se(elpd)
$\gamma_{\text{the}}, \gamma_{\text{all}}, \gamma_{\text{two}}$	-3072.2	57.6
$\gamma_{\text{the}}, \gamma_{\text{all}} = \gamma_{\text{two}}$	-3076.8	57.6
$\gamma_{\text{all}}, \gamma_{\text{the}} = \gamma_{\text{two}}$	-3081.4	57.6
$\gamma_{\text{all}} = \gamma_{\text{the}} = \gamma_{\text{two}}$	-3103.6	57.3
$\gamma_{\text{all}} = \gamma_{\text{the}}, \gamma_{\text{two}}$	-3104.3	57.4

Table 7.6: Estimated log predictive likelihood for models in Analysis 3 where $\beta_{\text{the}} = \beta_{\text{two}}$.

Parameters	elpd	se(elpd)
$\gamma_{\text{the}}, \gamma_{\text{all}}, \gamma_{\text{two}}$	-3071.8	57.5
$\gamma_{\text{the}}, \gamma_{\text{all}} = \gamma_{\text{two}}$	-3075.2	57.5
$\gamma_{\text{all}}, \gamma_{\text{the}} = \gamma_{\text{two}}$	-3082.4	57.6
$\gamma_{\text{all}} = \gamma_{\text{the}} = \gamma_{\text{two}}$	-3104.9	57.3
$\gamma_{\text{all}} = \gamma_{\text{the}}, \gamma_{\text{two}}$	-3105.2	57.4

Table 7.7: Estimated log predictive likelihood for models in Analysis 3 without constraints on β_{the} and β_{two} .

The availability of inverse-scope distributive readings for a determiner can be expected to be modulated by two factors: the availability of distributive readings with that determiner in general, and the propensity of the DP type to take inverse scope. Having established that the preference for distributive readings is to some extent determiner-specific even among plural DPs, we may now ask the same question about scope-taking behaviour: are any of *all*, *the*, and *two* more ready to take inverse scope than others?

We thus fitted models of the following form on the full data set minus the BOTH conditions. Again there were five models, varying by which, if any, of the three γ -parameters were constrained to be identical.

$$Y_{isd} \sim \text{bernoulli}(\text{logit}^{-1}(\pi_{isd}))$$

with: $\pi_{isd} = \alpha_d + u_{1s} + (\beta_d + u_{2s}) \cdot \text{DIST}_i + (\gamma_d + u_{3s}) \cdot \text{DIST}_i \cdot \text{INV}_i$

where $\alpha_{\text{all}} = \alpha_{\text{two}}$ (conclusion from Analysis 1). We fitted both models where $\beta_{\text{the}} = \beta_{\text{two}}$, as concluded from Analysis 2a, and models where this constraint was lifted. As can be seen from Tables 7.6 and 7.7, there was no appreciable difference between these two sets of models.

It is clear that idiosyncratic scope preferences do exist among the three determiners we investigated, as the model without idiosyncratic differences is notable worse than the next best model ($\Delta_{\text{elpd}} = 21.5$, $se(\Delta_{\text{elpd}}) = 6.9$). Beyond that, the picture that emerges is not entirely decisive, with possible pictures ranked as follows: *the* > *two* > *all* is somewhat better than *the* >

¹³Note that we cannot include a subject intercept in these models as it would absorb the effect in a model with a β -parameter set to 0 and thereby make it appear as good as one where the β -parameter is not set to 0.

all = two ($\Delta_{\text{elpd}} = -4.9$, $se(\Delta_{\text{elpd}}) = 3.5$), which in turn is somewhat better than *the = two > all* ($\Delta_{\text{elpd}} = -3.0$, $se(\Delta_{\text{elpd}}) = 6.9$).

It is worth noting that despite the fact that inverse-scope distributive readings are most available with *all*, it is actually the determiner with the greatest aversion to taking inverse scope: the availability of inverse-scope distributive readings is higher than with the other determiners purely because distributive readings in general are more readily available with *all*.

Part III

A methodological approach to disambiguation

In this part, I validate some methodological tools useful to investigate semantic processing and ambiguity resolution. In Chapter 8, I present a mouse-tracking study that seeks to identify the features in mouse trajectories that correspond to difference decision-making processes. Under the assumption that decision making and disambiguation have some core properties in common, the methodology developed in Chapter 8 is used in Chapter 9 to analyse data from plural ambiguities processing.

8 | Mouse tracking as a window into decision making^{*}

Maldonado, M., Dunbar, E. & Chemla, E. Mouse tracking as a window into decision making. Submitted.

Abstract

Mouse tracking promises to be an efficient method to investigate the dynamics of cognitive processes: it is easier to deploy than eye-tracking, and yet it is in principle much more fine-grained than looking at response times. We investigate its claimed benefits directly, asking how features of decision processes, and notably decision changes, may be captured in mouse movements. We ran two experiments, one in which we explicitly manipulate whether our stimuli trigger a flip in decision, and one in which we replicate more ecological, classical mouse tracking results on linguistic negation (Dale & Duran, 2011). We conclude, first, that spatial information (mouse path) is more important than temporal information (speed and acceleration) for detecting decision changes, and we offer a comparison of the sensitivity of various typical measures used in analyses of mouse tracking (area under the trajectory, direction-flips, and others). We do so using an ‘optimal’ analysis of our data (a linear discriminant analysis) explicitly trained to classify trajectories, and see what type of data (position, speed, acceleration) it capitalizes on. We quantify how its results compare with those based on more standard measures.

8.1 Introduction

In the past ten years, mouse tracking has become a popular method for studying the dynamics of cognitive processes in different domains, ranging from phonetic competition (Cranford & Moss, 2017; Spivey, Grosjean, & Knoblich, 2005), and syntactic, semantic and pragmatic processing (Dale & Duran, 2011; Farmer, Cargill, Hindy, Dale, & Spivey, 2007; Sauerland, Tamura, Koizumi, & Tomlinson, 2015; Tomlinson, Bailey, & Bott, 2013; Xiao & Yamauchi, 2014, 2017, among others), to social cognition (Freeman & Ambady, 2010; Freeman, Dale, & Farmer, 2011; Freeman & Johnson, 2016). While response times can reveal whether a decision process is fast or slow (Donders, 1969), and analyses of response time distributions can give insight into how the decision process unfolds (Ratcliff & McKoon, 2008; Usher & McClelland, 2001, among others), mouse movements promise a more direct window onto the dynamics of cognitive processes, under the assumption that motor responses are planned and executed in parallel to the decisions they reflect (Freeman & Ambady, 2010; Hehman, Stolier, & Freeman, 2014; Song & Nakayama, 2006, 2009; Spivey & Dale, 2006).

Concretely, if a response is entered by clicking on a button, one may measure the time needed to click on that button and use it as a reflection for the complexity of the decision, roughly. But depending on whether participants are decided from the start, hesitate, or undergo a radical change of decision, the path to that button may take different trajectories (see Figure 8.1,

^{*} The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n.313610 and was supported by ANR-10-IDEX-0001-02 PSL* and ANR-10-LABX-0087 IEC.

Wojnowicz et al., 2009). Accordingly, researchers have studied the shape and dynamics of mouse paths to document aspects of numerous types of decision processes. Dale and Duran's (2011) approach to negation processing is an example of this. Linguistic negation has been traditionally understood as an operator that reverses sentence truth conditions, inducing an extra "step," or "mental operation," in online processing (Wason, 1965; Wason & Johnson-Laird, 1972; see review in Tian & Breheny, 2016). Dale and Duran tracked mouse trajectories as participants performed a truth-value judgment task, where they had to verify the truth of general statements such as *Cars have (no) wings*. The authors found that mouse trajectories gave rise to more shifts towards the alternative response when evaluating negative than affirmative true sentences. This was interpreted as evidence for a "two-step" processing of negation, where truth conditions for the positive content are first derived and negated only as a second step.¹ To do so, one can extract several measures from the mouse paths (e.g., maximal deviation point, number of direction changes, etc.) and argue that the deviation of these measures from what they would be for an optimal, straight trajectory reflects the relevant decision change.

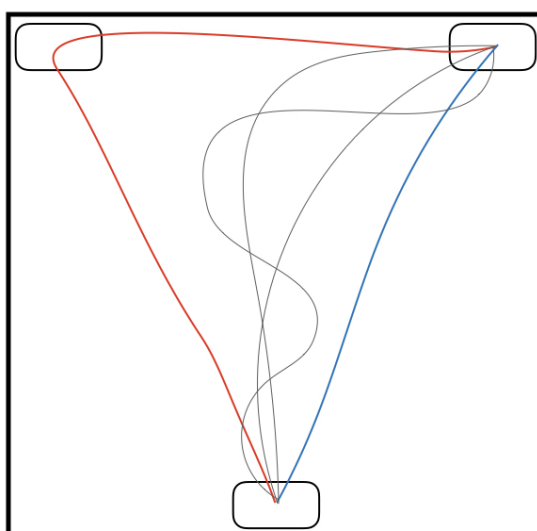


Figure 8.1: Shape of trajectories underlying distinct decision processes. One single cognitive process is expected to be mapped onto one smooth trajectory (blue line), whereas a change of mind would be reflected by a two-step path (red line). Intermediate cases are represented in gray.

Our goal is to explicitly document this method and the connection between cognition (decision making) and action (mouse trajectories): What in a decision process is reflected in mouse movements—decision changes, hesitations, or other properties?—and how—in changes in acceleration, changes in direction, or other aspects of the trajectory? We will tackle this question by asking what features of mouse trajectories distinguish *straightforward* decisions, based on a single initial commitment, and *switched* decisions, which involve a change of mind

¹Several studies have suggested that the positive argument plays an important role in negation processing (Kaup, Yaxley, Madden, Zwaan, & Lüdtke, 2007; Lüdtke, Friedrich, De Filippis, & Kaup, 2008, among others). This pattern of results, however, depends on the amount of contextual support given for the sentence: "two-step" negation processing seems to occur specifically for sentences presented out-of-the-blue, whereas no difference between negative and positive sentences arises when the right contextual support is provided (Nieuwland & Kuperberg, 2008; Tian, Breheny, & Ferguson, 2010). How to explain this pattern of results has been at the center of the debate in the negation processing literature (see Tian & Breheny, 2016 for review). We will not explore this here.

in the course of the process.

First, we present a validation experiment where we directly manipulate whether the stimuli trigger a flip in what the appropriate response is in the course of a trial. We show that the mouse paths do indeed reflect these changes (Section 8.2). An analysis of this data using linear discriminant analysis (henceforth, LDA), confirms that the two types of decision, straightforward and switched, can be distinguished objectively (Section 8.3). We then compare the performance of the LDA classifier to other traditionally used mouse tracking measures (Section 8.4). Finally, the LDA classifier trained on the validation data is further tested with new, more “ecological” data, obtained from a replication of Dale and Duran’s (2011) experiment on the processing of negation mentioned above (Section 8.5). If there is a change of decision triggered by negation, trajectories corresponding to negative trials should be classified together with trajectories underlying changes of decision in the validation experiment.

Data and code for all the analyses developed in this paper are provided at https://osf.io/rbx3m/?view_only=7d557aa8931c4a0886e7ce2442a77895.

8.2 Validation Experiment: presentation and qualitative analysis

Participants were asked to perform a two-alternative forced choice task. Each trial was triggered by clicking on a start button at the bottom of the screen. A frame surrounding the screen would then appear and the participants’ task was to indicate whether the frame was blue or red by clicking on the appropriate “blue” or “red” buttons at the top left or top right of the screen, respectively. On most trials, the color of the frame remained stable throughout the trials, but in crucial cases it changed during the trial. In the first case, the initial choice was the correct response (*straightforward* trials). In the second case, participants were forced to change their answer (*switched* trials). The switched trials are meant to mimic natural decision changes. We take these to be a reasonable stand-in for changes of decision, even though there are obvious differences: in natural changes of decision, alternative responses are weighted as the pieces of information are integrated, whereas in our experiment the sensory information changes in time. We return to the question of how ecological these decisions are in Section 8.5. The procedure is illustrated in Figure 8.2.

8.2.1 Participants

We recruited 54 participants (F=27) using Amazon Mechanical Turk. Two subjects were excluded from the analyses because they did not use a mouse to perform the experiment. All of them were compensated with 0.5 USD for their participation, which took approximately 5 minutes.

8.2.2 Design

Each trial instantiated one of two possible DECISION PATTERNS. In *straightforward* trials, the frame color remained stable, and the decision made at the beginning of the trial did not need to be revised. In *switched* trials, the color switched once (from red to blue or from blue to red) during the trial, forcing a revision of the initial choice. The change on *switched* trials was triggered by the cursor reaching a certain position on the *y*-axis, which could be at various relative heights (POINT OF CHANGE: early, at 40% of the screen, middle, at 70%, or late, at 90%). The design is schematized in Table 8.1.

To prevent participants from developing a strategy whereby they simply drag the cursor along the center line rather than moving the mouse toward their current choice of answer, the proportion of trials was adjusted so that there were a majority of 64 straightforward trials (32 repetitions per frame color), while there was only 24 switched trials (4 repetitions per final frame color and change point).

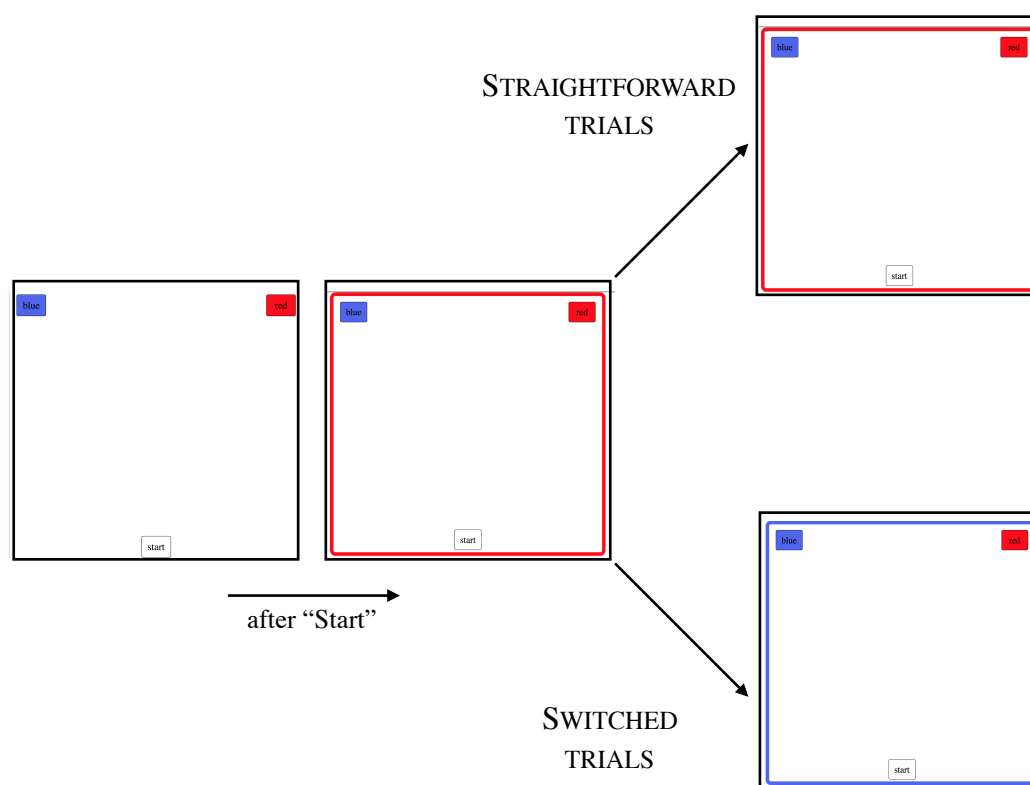


Figure 8.2: Procedure in Validation Experiment. Subjects were instructed to click the “start” button in order to see the colored frame. Response boxes were on the top left or top right. Depending on the trial condition, the frame color either did, or did not, change (once) during the trial.

DECISION PATTERN	FRAME COLOR	POINT OF CHANGE
Straightforward	Blue	<i>never</i>
	Red	
Switched	Blue → Red	early (y=40%)
	Red → Blue	middle (y=70%)
		late (y=90%)

Table 8.1: Design in Validation Experiment

8.2.3 Interface

The web interface was programmed using JavaScript. Mouse movements triggered the extraction of (x, y) -pixel coordinates (there was thus no constant sample rate). Three buttons were displayed during the experiment (“start” and response buttons). The “start” button was placed at the bottom center of the screen. The two response boxes were located at the top left (“blue”) and top right (“red”) corners. On each trial, between start-clicks and response-clicks, mouse movements triggered the recording of the (x, y) -pixel coordinates of the cursor together with the time.

8.2.4 Data treatment

To allow comparisons between participants, the (x, y) -coordinates were normalized according to participants’ window size: the center of the start button was mapped onto $(0, 0)$ point, the “blue” button onto $(-1, 1)$ and the “red” button onto $(1, 1)$. Variations both in response times and in the sensitivity and sampling rate of our participants’ input devices imply that different

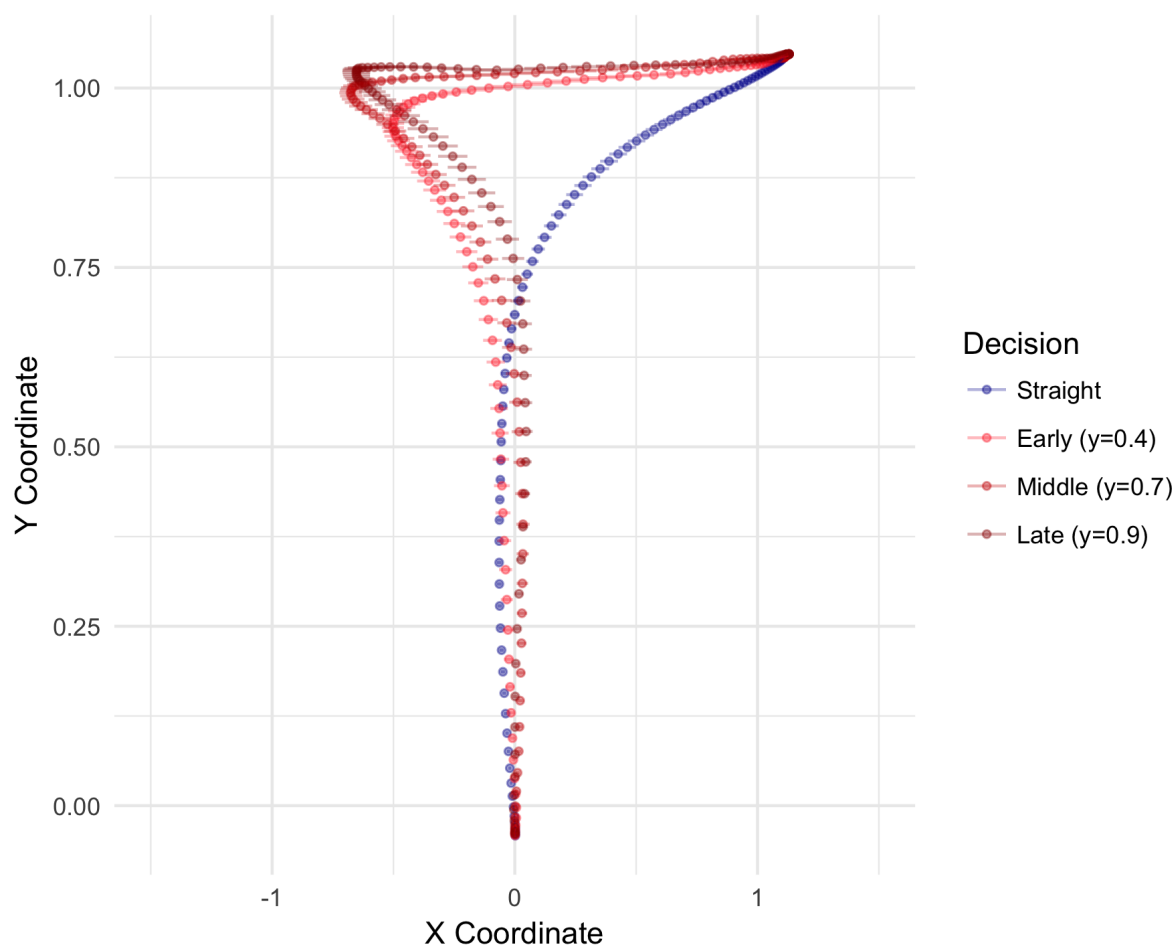


Figure 8.3: Mean trajectories for different decision patterns in the validation experiment. Error bars represent the standard error of x -coordinates.

trials have different numbers of (x, y) positions per trial, making comparisons difficult. We therefore normalized the time course into 101 proportional time steps by linear interpolation. That is, we reduced all time points to 101 equally distant time steps, including the first and the last positions.

8.2.5 Overall performance

Inaccurate responses (4% of the data) were removed from the analyses. Mean trajectories for each DECISION PATTERN and POINT OF CHANGE are illustrated in Figure 8.3. These trajectories suggest that participants made a decision as soon as they were presented with the color frame, and revised this decision if needed. When they were forced to change their choice, this switch was reflected in mouse trajectories.

8.3 Validation Experiment: Classifying decision processes with LDA

Different decisions (that is, DECISION PATTERNS) have a different impact on mouse trajectories (Figure 8.3). To identify the features characteristic of each class (*switched* vs. *straightforward*), we use a linear discriminant analysis for classification.

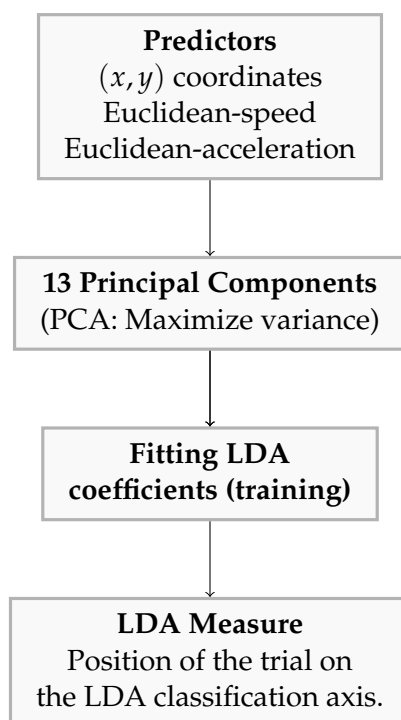


Figure 8.4: Diagram of classification procedure.

8.3.1 Description of LDA classifier

The LDA is a supervised algorithm that finds a linear function of the predictors onto a single real number, such that zero represents the midpoint between the two classes to be learned, and the separation between the two classes is maximal. This linear combination of predictors can thus be used to form a decision rule to classify objects of one class (negative) or the other (positive).

The two classes here were the multi-dimensional data coming from *switched* and *straight-forward* trials. The dimensions taken into account were: all the x, y coordinates, the Euclidean-distance based velocity and the Euclidean-distance based acceleration (both of which are non-linear with respect to the original (x, y) coordinates). The coordinates provide absolute spatiotemporal information about where the cursor was at what point, and velocity and acceleration provide information about how it arrived there. To avoid collinearity (which causes problems for LDA), we applied a principal component analysis (PCA) to identify 13 principal components for these predictors, and fitted and applied the LDA to these principal components. We thus obtained an *LDA measure* for each trial, the single number giving the position of the trial on the LDA classification axis. The procedure is schematized in Figure 8.4.

8.3.2 Performance of the LDA classifier

Figure 8.5 illustrates the result of applying the procedure in Figure 8.4 to the trajectories. To evaluate the overall performance of the classifier, we calculated the area under the ROC curve (AUC), a standard method for evaluating classifiers (Hastie, Tibshirani, & Friedman, 2009). Intuitively, the AUC gives the degree to which the histograms resulting from the classifier's continuous output (for example, Figure 8.5) are non-overlapping in the correct direction (in this case, *switched* more systematically in the positive direction on the classification axis than *straightforward*).

To properly evaluate the classifier's performance at separating trials following the distribution in the experiments, the AUC measure was cross-validated. That is, calibration data were

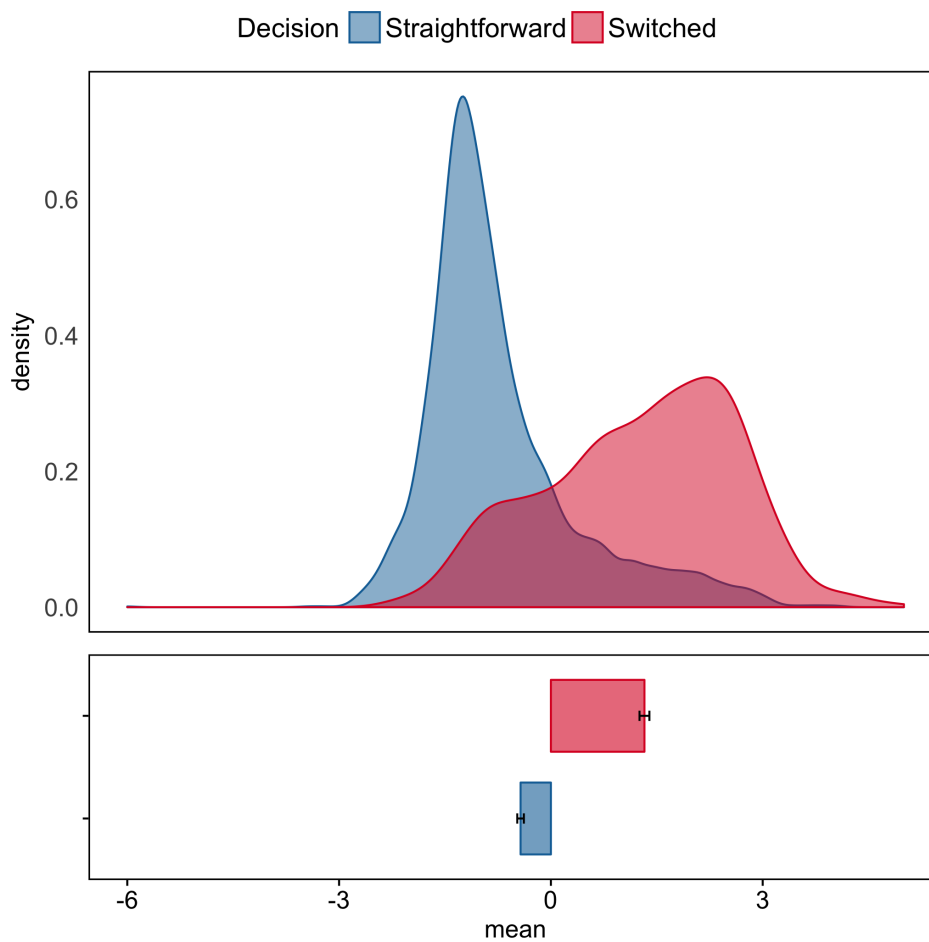


Figure 8.5: Distribution and mean LDA-based measure for each class. Classifier performance when applied to the whole validation data set. Error bars represent the mean standard error.

partitioned into 10 bins that kept the proportion of *straightforward* and *switched* trajectories constant (75/25 proportion). For each bin, we took the complementary set of data (the remaining 90%) to train the classifier. The data contained in the bin were used as a test set to diagnose the classifier performance. We thus obtained one AUC score for each of the ten test bins. The performance of the LDA classifier was compared to a baseline, equivalent to the worst possible outcome, and a topline, which was what we would expect from a LDA under the best possible conditions. For the baseline, we used a random classifier that assigned labels by sampling from a beta distribution centered at the probability of straightforward trials; the topline was computed by testing and training the original LDA classifier on the same set of data. The mean AUC values for the LDA, the baseline and the topline in each bin are given in Figure 8.9a.

To assess whether the performance of the LDA classifier was statistically different from baseline (or topline) performance, we tested the groups of ten scores with regard to how likely it would be to obtain the attested differences in scores under the null hypothesis that the LDA classifier performance was the same as baseline (or topline) performance. The difference in the mean AUC between each of these two pairs of classifiers was calculated as a test statistic. The sampling distribution under the null hypothesis was estimated by randomly shuffling the labels indicating which classifier the score came from.

In Table 8.2a, we report the results of performing a one-tailed test on the mean AUC differences. As expected, our original LDA is significantly better than a random classifier at categorizing trajectories into *straightforward* and *switched*. Conversely, there is no significant difference between the performance of our LDA and the topline; the classifier's performance is not significantly different from the best an LDA could possibly give on this data.

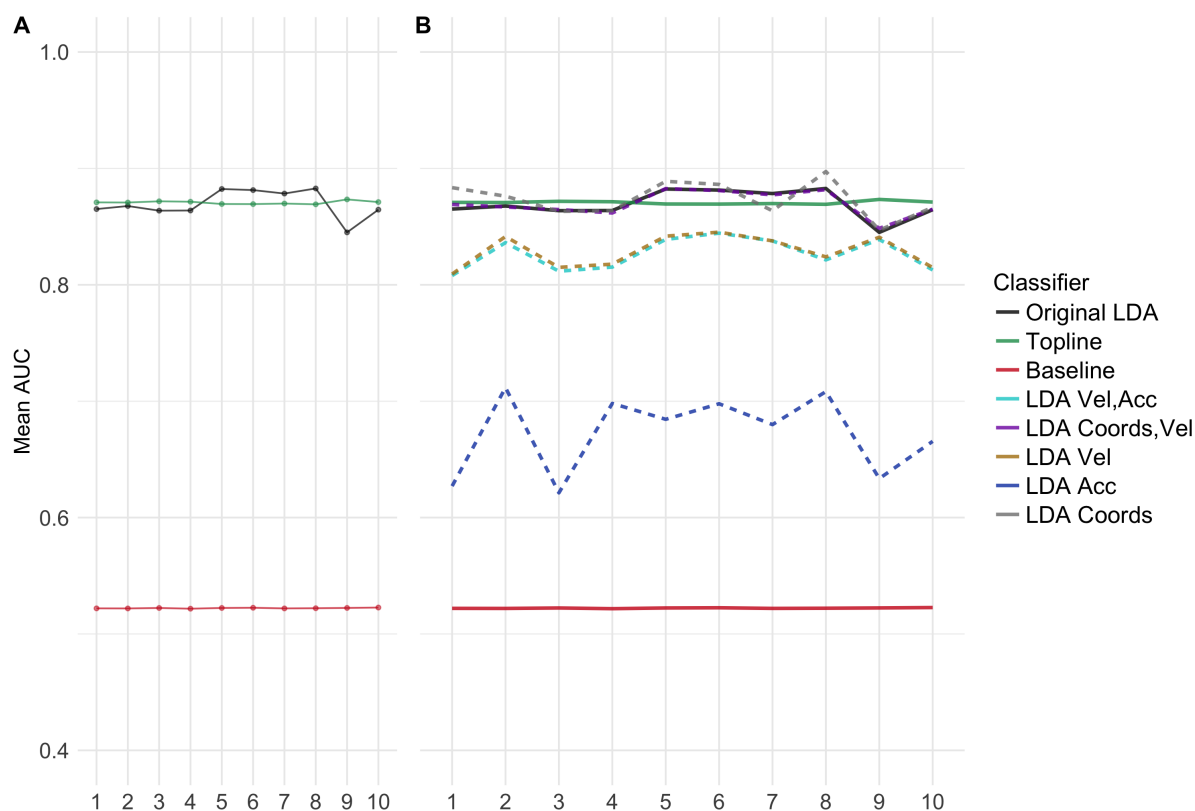


Figure 8.6: Mean Area Under the ROC Curve values obtained from cross-validation. A. Cross-validation on 10 bins for original LDA, baseline and topline. B. Comparison with values obtained for five additional classifiers obtained by subsetting the original set of predictors.

	ORIGINAL LDA (coords, speed, acc)	(a)		(b)				
		BASELINE	TOPLINE	LDA WITH DIFFERENT PREDICTORS				
				coords, vel	acc	coords	vel	acc
AUC (mean)	.87	.52	.87	.87	.83	.87	.82	.67
Mean Difference	–	.35	-.002	-.0004	.04	-.006	.04	.2
<i>p</i> value	–	<.001	0.58	.5	<.001	.68	<.001	<.001

Table 8.2: Cross-validation results for the LDA classifier. The performance of the LDA was compared to the one of (a) baseline and topline classifiers and (b) LDA classifiers with different predictors.

8.3.3 Meaningful features and optimal predictors

Our original LDA classifier takes as predictors both absolute and relative spatio-temporal features (coordinates, speed, and acceleration). Some of these features, however, might not be relevant for the classification. By comparing classifiers trained with different predictors, we gather information about which features of mouse trajectories are most relevant to decision processes.

We trained five additional LDA classifiers obtained by subsetting the three original LDA predictors. If both absolute and relative features are required to predict the decision type, we would expect our “full” original LDA classifier to be better than any other classifier that takes only a subset of these original predictors. The performance of these additional classifiers was diagnosed in the same way as before, by computing the AUC for each of the 10 test bins. Figure 8.9b illustrates the mean AUC values for each of these classifiers, together with the original LDA, the baseline and the topline. Pairwise comparisons with the original LDA were done by testing whether the observed mean differences would be expected under the null hypothesis of no difference in performance between classifiers. Table 8.2b summarises the comparisons between each of these classifiers and our original LDA.

The original LDA does not significantly differ from other LDA classifiers that contain the coordinates among their predictors, suggesting that the distinction between *straightforward* and *switched* decisions might be solely explained by the information contained in the (x, y) coordinates. Conversely, the original LDA is significantly better than classifiers that use only speed and acceleration as predictors. These comparisons therefore reveal that, for classifying our validation data, absolute spatio-temporal features ((x, y) coordinates) are generally better predictors than relative features (speed and acceleration). That is, it seems to be more relevant to know where the mouse pointer was at a given time than to know how it got there.

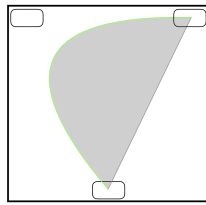
We caution that effects of *true* decisions, rather than the simulated decisions tested here, may indeed have an impact on speed and acceleration. It has been suggested that speed and acceleration components can capture the level of commitment towards the response, such that a change of decision (*switched* trajectories) might have associated with it a specific speed/acceleration pattern (Hehman et al., 2014). This is not visible, however, in our data.

8.4 Validation Experiment: LDA versus traditional mouse tracking analyses

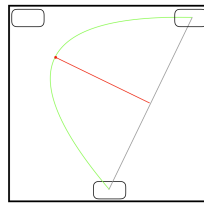
The LDA classifier derives a solution to the problem of separating two kinds of mouse trajectories that is in a certain sense optimal. Previous studies have used alternative techniques to analyze mouse trajectories. In what follows, we compare the performance of our LDA to other measures commonly used in mouse tracking studies. We focus on measures that assess the spa-

tial disorder in trajectories, typically taken to be indicative of unpredictability and complexity in response dynamics (Hehman et al., 2014).

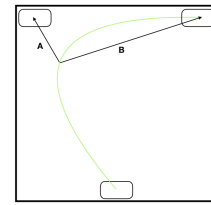
Two of the most commonly used methods of mouse tracking **spatial analysis** are the *Area under the trajectory* and the *Maximal deviation* (henceforth, AUT and MD respectively) (see Freeman & Ambady, 2010). The AUT is the geometric area between the observed trajectory and an idealized straight-line trajectory drawn from the start to the end points, whereas the MD is the point that maximizes the perpendicular distance between this ideal trajectory and the observed path (Figure 8.7). For both measures, higher values are associated with higher trajectory deviation towards the alternative; values close to or below zero suggest a trajectory close to ideal. Another frequently used measure counts the number of times a trajectory crosses the x -axis (horizontal flips, Dale & Duran, 2011, as illustrated in Figure 8.7).



(a) Area Under the Trajectory



(b) Maximal Deviation



(c) Maximal Log Ratio

(d) X-coordinate Flips

$$\sum H[(x_t - x_{t-1})(x_{t-1} - x_{t-2})]$$

(e) Acceleration flips

$$(\sum H[(a_t - a_{t-1})(a_{t-1} - a_{t-2})]) - 1$$

Figure 8.7: Description of commonly used mouse tracking measures.

While all these measures aim to evaluate the degree of complexity of the path, they may fail to distinguish paths straight to the correct answer from “two step” (deviation to the alternative) and from “uncertain” (centered on the middle of the screen) trajectories.² To assess more directly whether mouse trajectories have a meaningful deviation towards the alternative, the distance to both target and alternative responses should be taken into account. For instance, the *ratio of the target distance to the alternative distance* can be calculated for each (x, y) position. While ratio values closer to one suggest a position near the middle, higher values indicate a deviation towards the alternative response.

AUT, MD, x -coordinate flips, and the point that maximizes the log distance ratio (henceforth Maximal Log Ratio) were calculated for the validation data. Following Dale and Duran (2011, and other studies on error corrections), we also analyzed the *acceleration component* (AC) as a function of the number of changes in acceleration. Since stronger competition between alternative responses is typically translated into steeper acceleration peaks, changes in acceleration can be interpreted as decision points (Hehman et al., 2014). Figure 8.8 illustrates the distribution and mean values for each decision pattern.

The same cross-validation procedure described in the previous section was used to diagnose the performance of each of these measures.³ The mean AUC values for each of these measures are illustrated in Figure 8.9. Table 8.3 summarizes the result of comparing the LDA performance to each of the alternative measures.

Overall, these comparisons reveal that the LDA trained on the validation data is significantly

²A late medium-size deviation towards the alternative might underlie a “two-step” decision, whereas an early, but large, deviation towards the alternative might very well be considered noise. Measures such as the AUT might not be able to make a distinction between these.

³Note that these measures do not need training; we simply applied the measure to the same ten test subsets as before to make the results comparable.

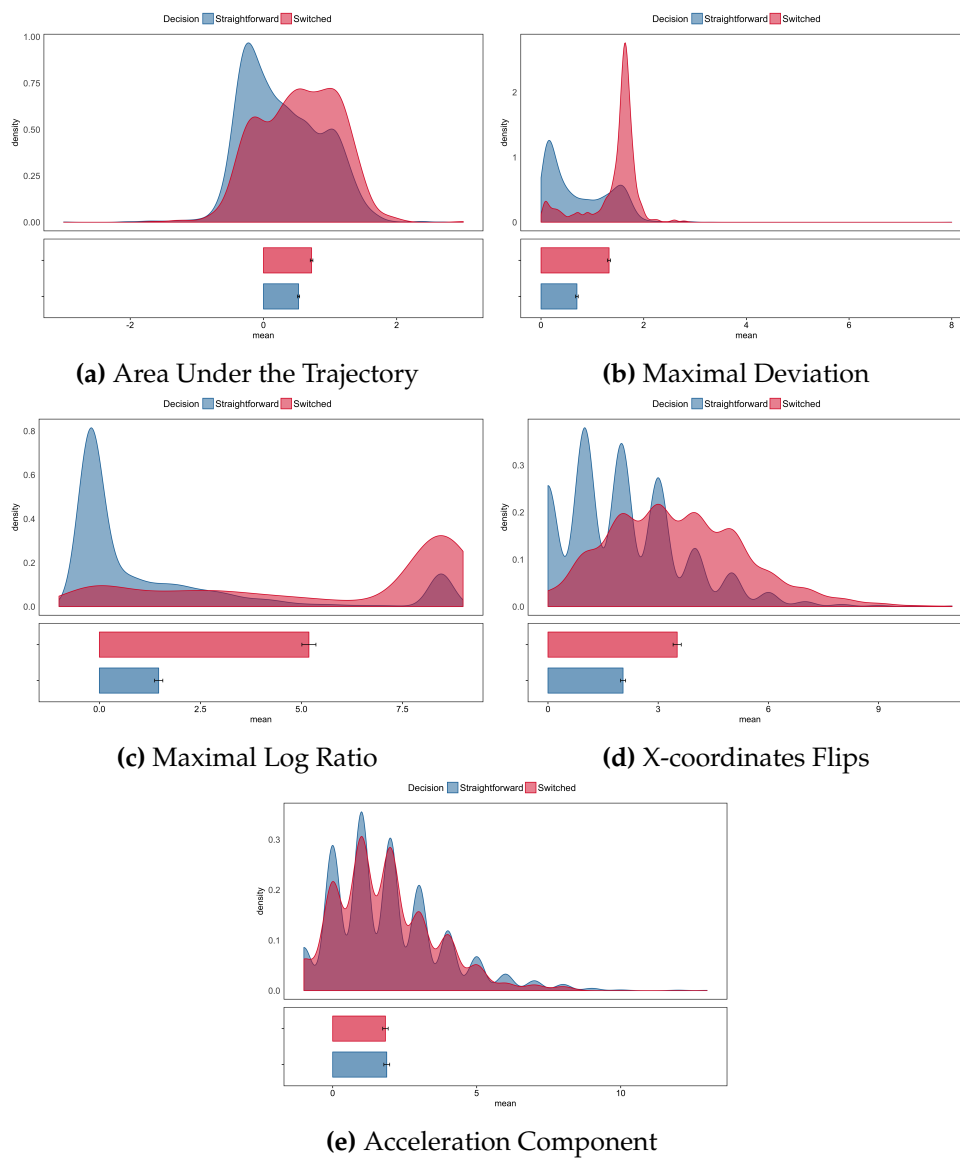


Figure 8.8: Distribution and means obtained from applying different mouse tracking measures to validation data. Error bars represent the mean standard error.

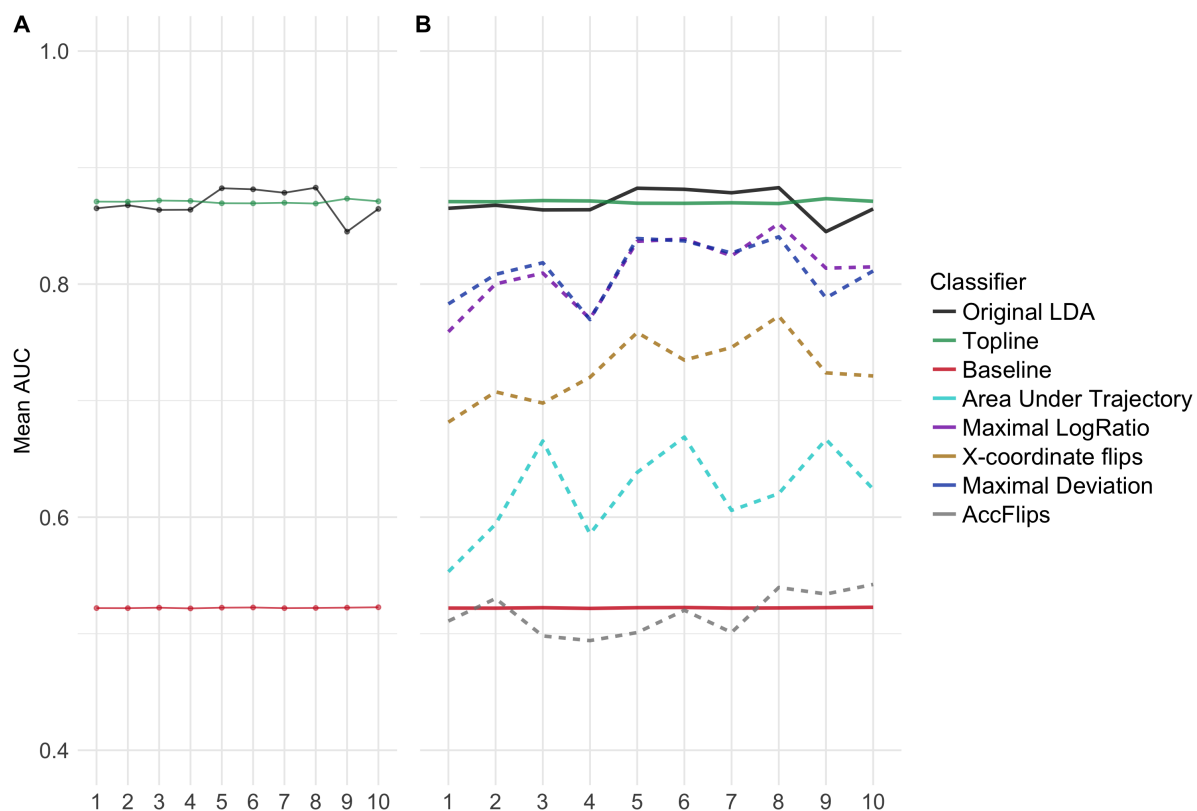


Figure 8.9: Mean Area Under the ROC Curve values obtained from cross-validation. A. Cross-validation on 10 bins for original LDA, baseline and topline. B. Comparison with values obtained for other commonly used mouse tracking measures.

	ORIGINAL LDA	AUT	MD	MAXIMAL LOGRA- TIO	X- COORD. FLIPS	AC
AUC (mean)	.87	.62	.81	.81	.73	.53
Mean Difference	–	.24	.06	.06	.14	.34
<i>p</i> value	–	<.001	<.001	<.001	<.001	<.001

Table 8.3: Cross-validation results for the LDA classifier. The performance of LDA was compared to each of five commonly used measures in mouse tracking studies.

better at classifying this type of decisions than other commonly used measures. The difference with the classifier is in all the cases significant. Mean AUC values suggest that MD and the Maximal Log Ratio are better at distinguishing decision processes than the other alternative measures. These two measures are the only ones calculated based on coordinates, and therefore give more importance to spatio-temporal information than the others. In other words, the MD and the Maximal Log Ratio are not only sensitive to whether or not there was a deviation from the ideal trajectory (as the other measures), but weight this deviation as a function of the moment at which it occurred, assigning higher values to late deviations. This information seems to be essential for the classification, as observed in Section 8.3.

Finally, we previously observed that velocity and acceleration were not helpful predictors for the LDA classifier. Indeed, the performance of the Acceleration Component overlaps here with that of the Baseline, suggesting that this type of information is not helpful.

8.5 Extension to linguistic data

So far, we have shown that (i) a rough manipulation of decision making processes has a direct impact on mouse trajectories; (ii) an LDA using absolute temporal information is enough to accurately distinguish these decision patterns; and (iii) this LDA does a better classification than other traditional mouse tracking measures. Can our LDA help characterize more complex decision processes, such as the ones involved in sentence verification tasks?

To address this question, we test our classifier on data obtained from a replication of Dale and Duran’s experiment (2011). This experiment found differences in the processing of true positive and negative sentences when people performed a truth-value judgment task. These results were interpreted as indicating that negation gives rise to an abrupt shift in cognitive dynamics (an unconscious change of decision). If this is indeed the case, we would expect mouse trajectories corresponding to the verification of negative sentences to pattern with *switched* trajectories from the validation experiment. This pattern of results would provide additional support to the hypothesis that, at least in out-of-the-blue contexts, processing negation does involve two steps, in which the positive value is initially derived and negated only as a second step. On the other hand, if negation does not involve a change in decision—or if subjects’ behavior in the validation experiment is simply too different from natural changes of decision—then the LDA measure trained on validation data will not reveal systematic differences between positive and negative sentences.

8.5.1 Experiment

Participants had to perform a truth-value judgment task in which they had to decide whether a sentence (for example, *Cars have wheels*) is true or false, based on common world knowledge. Each sentence could either be a negated form or a non-negated form, and could either be a

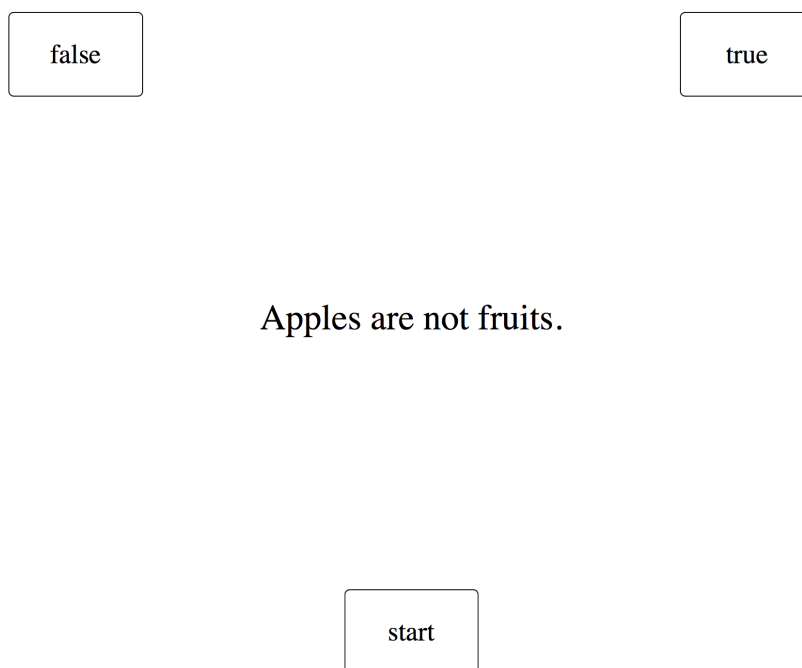


Figure 8.10: Illustration of a trial in the replication of Dale & Duran.

true or a false statement. Unlike Dale and Duran’s experiment, the complete statement was presented in the middle of the screen after participants pressed “start” (that is, no self-paced reading). The response buttons appeared at the top left or top right corners of the screen, as in our validation experiment. Materials and design are exemplified in Table 8.4.

Participants

53 English native speakers ($F=29$) were tested using Amazon Mechanical Turk. They were compensated for their participation (1 USD). The experiment lasted approximately 10 minutes.

Design

The experimental design consisted of two fully crossed factors: TRUTH VALUE (true, false) and POLARITY (negative, positive). We had a total of 4 conditions, and each participant saw 4 instances of each condition (16 sentences).

Truth value	Polarity	Example
True	Positive	Cars have wheels.
	Negative	Cars have no wings.
False	Positive	Cars have wings.
	Negative	Cars have no wheels.

Table 8.4: Design of Dale and Duran’s replication.

Interface and data treatment

The interface and data treatment were the same as those used in the validation experiment. The time course of mouse trajectories was normalized into 101 time steps.

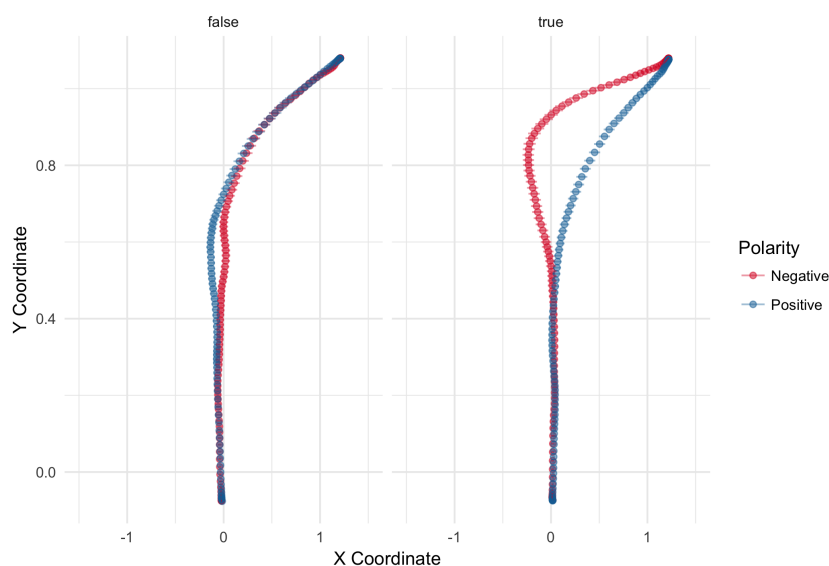


Figure 8.11: Mean trajectories for accurate trials

8.5.2 Results and discussion

Replicating Dale and Duran (2011)

All participants responded correctly more than 75% of the time. No participant was discarded based on accuracy. Only accurate trials were analyzed. Figure 8.11 illustrates mean trajectories for the four conditions.

To assess whether we replicate Dale and Duran’s results, we calculated the x -coordinate flips (see Section 8.4) and analyzed them with a linear mixed-effects model, taking TRUTH, POLARITY and their interaction as predictors. We included random intercepts per subject and a random slope with the interaction of both factors. P -values were obtained by comparing the omnibus model to a reduced model where the relevant factor was removed. This is the analysis done by Dale and Duran. Unlike Dale and Duran, we did not perform statistical analyses based on the acceleration component, since this quantitative measure was unable to distinguish mouse trajectories underlying different decision patterns in the validation experiment.

The model for x -coordinate flips revealed a main effect of POLARITY, such that negation increased the number of flips by an estimated of 0.76 ($\chi^2 = 21.7$; $p < .001$), and a significant interaction TRUTH \times POLARITY ($\chi^2 = 24.7$; $p < .001$), such that the difference between negative and positive sentences is bigger for the true than for the false statements. There was no significant effect of TRUTH ($\chi^2 < 1$; $p = .5$). Table 8.5 summarizes our and of Dale and Duran’s results.

Condition	x -flips	x -flips in D&D
T/no negation	2.22	1.13
T/negation	3.67	1.71
F/no negation	2.82	1.24
F/negation	2.9	1.34
Estimate Polarity	.76	.35
Estimate Truth	.07	.13
Estimate Truth \times Polarity	1.35	.47

Table 8.5: Mean and effect estimates for Dale & Duran original experiment and our replication.

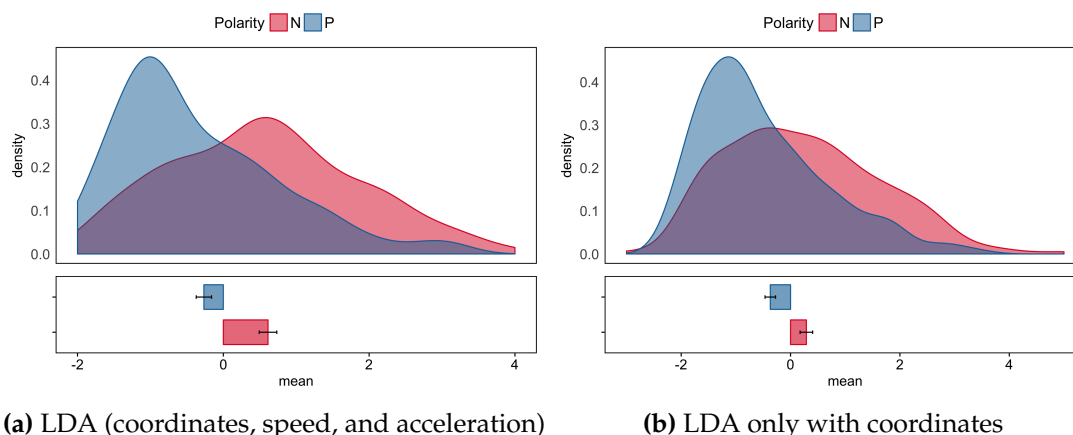


Figure 8.12: Two LDA classifiers applied to *true* trials (negative vs. affirmative). Error bars represent the mean standard error.

We seem to replicate Dale and Duran’s findings: verifying true negated sentences produces less straightforward trajectories than true positive sentences. The values obtained in the two experiments are slightly different; our results present a higher range of values (see Table 8.5). In our experiments, the mouse position was not sampled at a fixed rate, creating additional noise which could be responsible for the range difference.

Classifier performance

How well does our LDA classify new trajectories underlain by cognitive processes that might, or might not, involve different decision patterns across conditions? Two different LDA classifiers, trained with the data from the validation experiment, were applied to the new experimental data. The first classifier was our original LDA, which had as predictors (x, y) coordinates as well as distance-based velocity and acceleration. The second LDA had only (x, y) coordinates as predictors. Validation results (see Section 8.3) suggest that the simpler model, which only relies on absolute information, might be sufficient to classify the two basic kinds of decision-making processes. That is to say, the simple model fits the data just as well as a more complex model, and can be interpreted more straightforwardly.

The relevant difference in processing between positive and negative sentences is expected to arise specifically for *true* statements. Consequently, we analyze the performance of both classifiers when applied to true trials. Figure 8.12 illustrates the distribution and means of the resulting *LDA measure*.

To assess how well these classifiers separate positive from negative trials, we bootstrapped 1000 new samples of various different sample sizes from the data from the replication experiment and calculated the area under the ROC curve for the classification of each one. Figure 8.13A shows the mean AUC values obtained after applying the classification procedure across these various samples of different sizes. The values are generally lower than the ones obtained in the validation experiment. This could be due to the fact that the tasks were different; or it could simply reveal idiosyncrasies of the original validation experiment data, or of this replication experiment.

Might the observed performance be expected, even if negative and positive trials were actually not systematically different? Are these AUC values significantly different from the ones one would have obtained from applying the LDA to a set of data where there is no difference between experimental conditions? We calculated the AUC values for a set of data where experimental labels (positive, negative) were scrambled (once per sample). The distribution of AUC values under this null hypothesis was compared to the performance observed for the original set of data. Figure 8.13B illustrates the separability of the two classifications for each

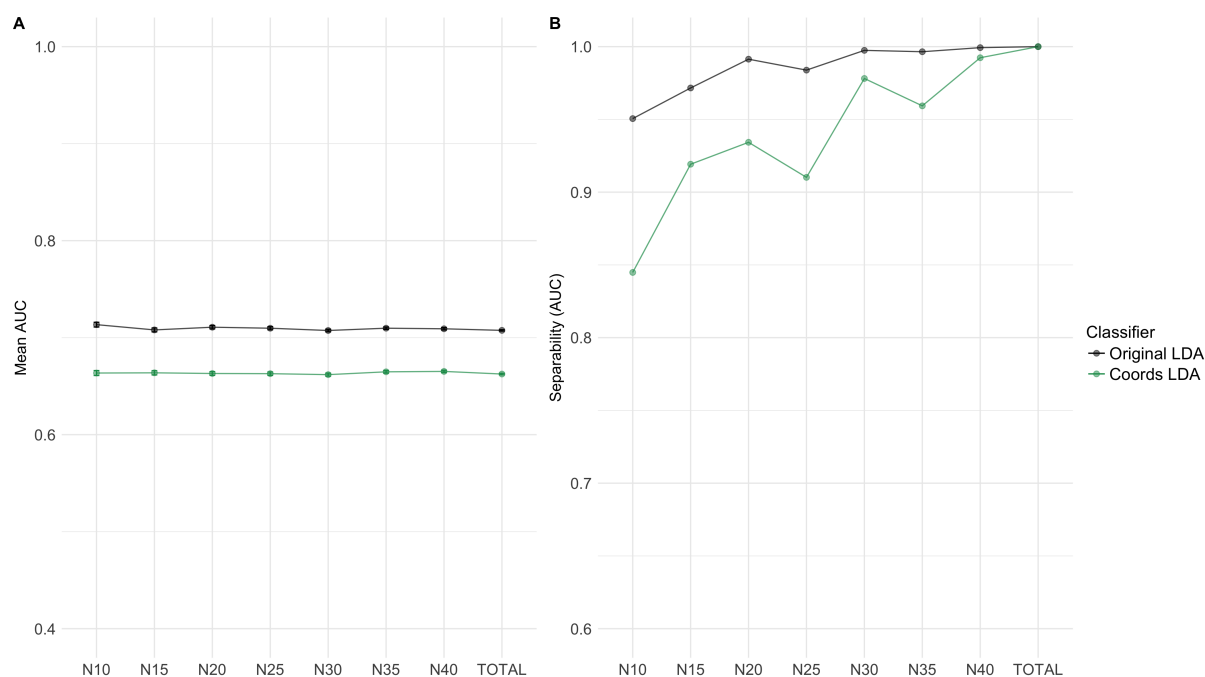


Figure 8.13: Performance of LDA classifiers. A. Mean AUC values over bootstrapped data (iterations=1000) for different sample sizes; B. Difference of classifier performance when applied to scrambled vs. original set of data.

sample size.

The LDA classifier trained with validation data seems to make a distinction between experimental conditions. This finding suggests that the contrast between negative and positive trials is similar to the contrast in the validation experiment. The fact that negation has similar properties to *switched* decisions indicates that verifying negative sentences might give rise to a change of decision, as proposed by Dale & Duran (2011), among others. However, while mouse trajectories corresponding to negative and to *switched* trials do share basic properties, they seem to differ on how they are placed on the “change of decision” spectrum: they occupy different parts of the decision-based LDA continuum (compare Figure 8.5 and Figure 8.12). This is not surprising, given that we are dealing with different cognitive processes—simulated decisions versus sentence verification—but, as discussed above, could easily also be an idiosyncrasy of these two data sets.

Finally, while the classifiers’ comparison in Figure 8.9 indicated that *relative* spatio-temporal features, such as acceleration and speed, were not essential for the classification of simple decisions, these features do seem to play a role in the classification of sentence verification data. Indeed, Figure 8.13 reveals that the *full* classifier—which takes all features as predictors—gives a better separation between the two experimental conditions than the simplified one.

8.5.3 Other mouse tracking measures

Does the difference in performance between the LDA and other mouse tracking measures remain when these are applied to the new experimental data? Figure 8.14 illustrates the distribution of each measure. The question of whether different measures differ in their ability to separate the experimental conditions was addressed by applying the same procedure as before: we calculated the mean area under the ROC curve for different sample sizes (see Figure 8.15A), and contrasted these values against a null hypothesis of no difference between experimental conditions (see Figure 8.15B).

The results in Figure 8.15A suggest that most measures perform less well here than on the

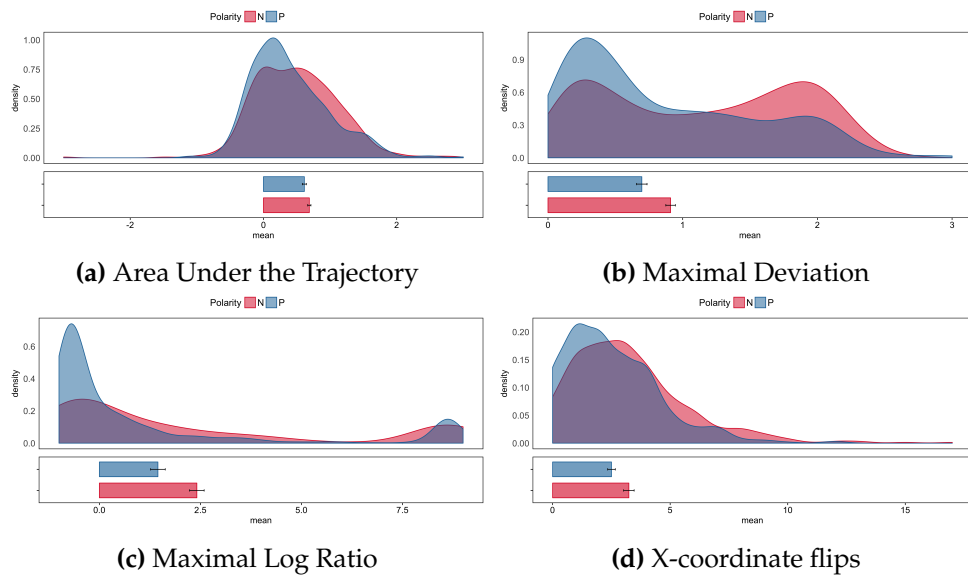


Figure 8.14: Distribution and means of negative and positive *true* trials obtained from applying different mouse tracking measures to negation data. Error bars represent the mean standard error.

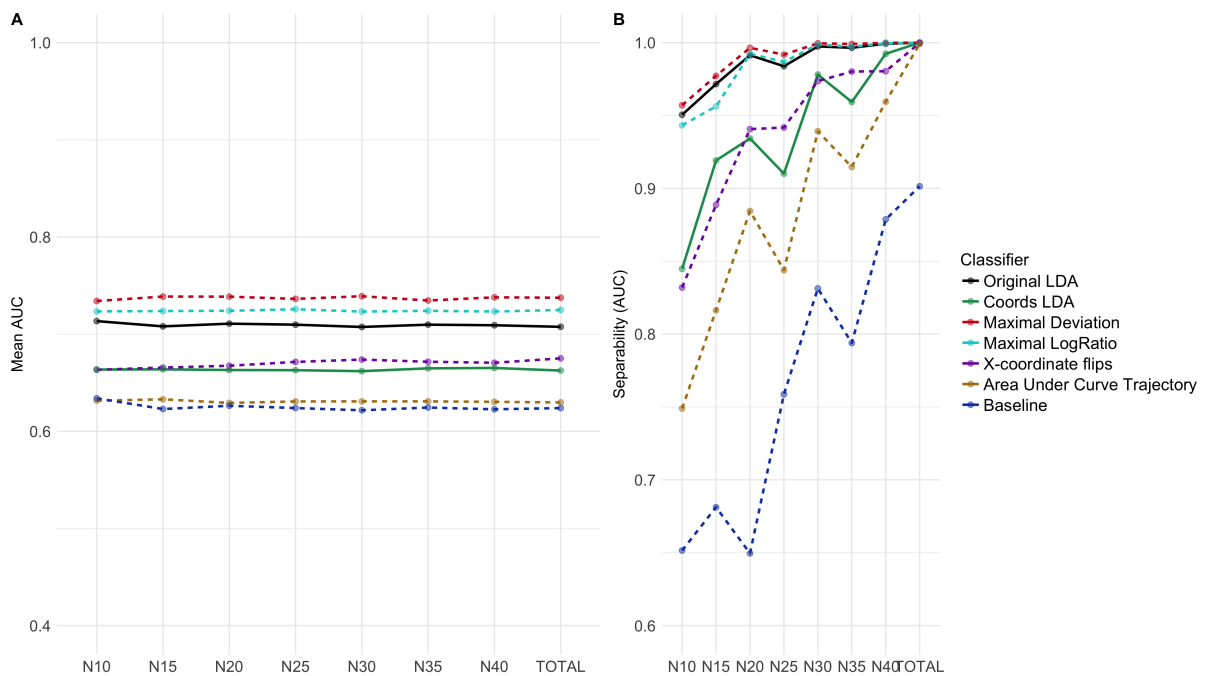


Figure 8.15: Performance of other mouse tracking measures. A. Mean AUC values over bootstrapped data (iterations=1000) for different sample sizes; B. Difference of measure performance when applied to scrambled vs. original set of data.

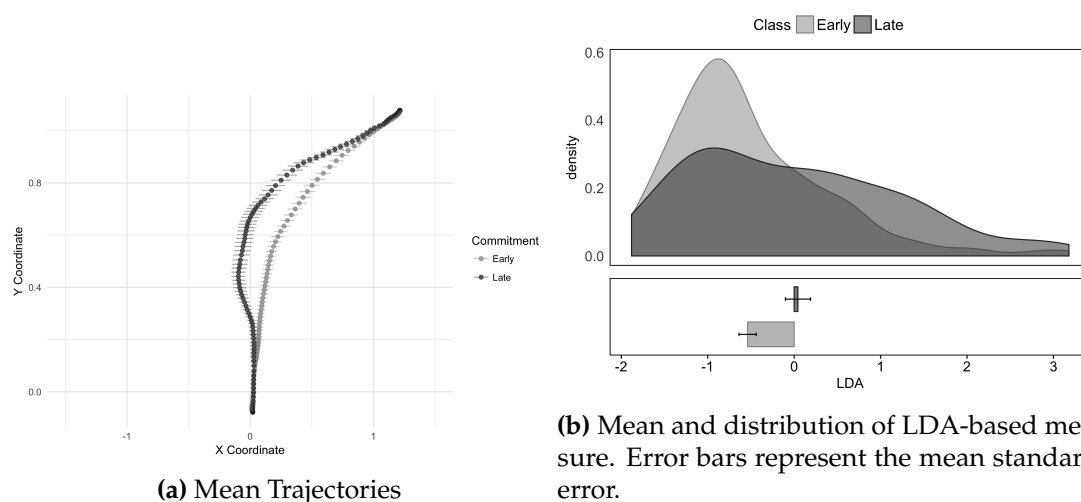


Figure 8.16: Analyses performed on Baseline data set (early vs. late decision).

validation data (cf. Figure 8.9). Since a decrease in performance is attested across the board and not only for the classifiers trained with validation data, this difference must be driven by properties of the new data set. The sentence verification data might be more variable, such that both negative and positive trials may underlie instances of different decision processes.

The LDA classifier seems here to be roughly as powerful as other traditional mouse tracking measures, such as the Maximal Deviation and the Maximal Log Ratio. In contrast with the validation results, this opens the possibility of using these alternative measures to analyze mouse tracking data from sentence verification tasks. The classifier is still a better choice from a conceptual point of view, as it does not make any specific assumptions about how the change of decision should be reflected by mouse trajectories beyond the observed.

Baseline

A linear classifier trained on simulated decisions can separate the two experimental conditions of the replication of a previous study by Dale & Duran's. We have interpreted this result as suggesting that the key features being extracted reflect two different decision processes. It could instead be argued that the classification is not based on properties related to decision processes, but on some other feature of mouse paths which happen to be partially shared between conditions in both experiments. For example, the LDA might be sensitive not to decision shift but to differences in cognitive cost, something both experiments might have in common.

To disentangle these possibilities, we asked how the classifier trained on simulated decisions classifies trajectories that have different shapes but ought not to be related to differing decision processes. We constructed a set of baseline data which contained only positive trials from the replication of the experiment by Dale and Duran. The trials were classified as to whether their response time was above or below the subject mean. We reasoned that shorter response times would correspond to early commitment towards the response, whereas longer response times would reflect a late commitment. As illustrated by in Figure 8.16a, the two classes in the baseline data have slightly different trajectory shapes. Importantly, however, nothing about this split implies that these shapes correspond to a change of decision. Thus, the classifier trained on *straightforward* versus *switched* trials was expected to perform poorly.

The distribution of the LDA measure after testing the classifier on the new data set is shown in Figure 8.16b. The performance was evaluated following the same procedure applied above (see blue line in Figure 8.15).

The classification on *early* versus *late* categories is less accurate than the one performed

to separate negative and positive trials. Differences in trajectories that are not due to the experimental manipulation are poorly captured by the LDA measure: even trajectories that have look similar to *switched* and *negation* trials are not taken to be underlying a change of decision. Thus, despite the differences between the experimental conditions in the validation experiment and in the replication experiment, the similarities appear to be more than accidental.

8.6 Conclusion

We investigated the correspondence between some types of decision processes and mouse movements. By manipulating whether a stimulus triggered, or did not trigger, a change of decision, we showed directly, for the first time, how mouse trajectories are impacted by decision changes: a forced switch in decision has an impact on mouse movements, which is for the most part observable in the spatial information (the path), and not so much in the timing of the trajectory.

We trained a classifier on the mouse trajectories underlying these simulated decisions to predict whether or not a given trial involved a decision shift. The classifier is freely available online. It accurately classifies not only paths corresponding to quasi-decisions, but also paths underlying a more complex process: the verification of negative sentences. Our results replicate previous findings but, more importantly, the LDA classifier performs at least as well as the best of the other commonly used mouse tracking measures. It also has the unique advantage of not relying on assumptions about what change-of-decision trajectories should look like. We also established the Maximal Deviation and Maximal LogRatio measures as comparable alternatives to the LDA analysis.

9 | Mouse tracking techniques applied to plural ambiguities

The data presented in this chapter were collected as part of my masters' project in 2015, before developing the data analysis tools presented in Chapter 8. In the light of the results reported in the previous chapter, I decided to reanalyse these data. For the sake of exposition, some aspects of the design were simplified here. The reader is referred to my masters' thesis for a complete description of the experiments.

9.1 Introduction

In Chapter 8, we have shown that a mouse tracking method can be used to capture differences in decision processes of various complexity. We have done that by relying on an 'optimal' analysis of mouse tracking data: a linear discriminant analysis (LDA) specifically trained to classify mouse paths based on whether the underlying decision process involved or not a change of decision.

In this chapter, I take one step further: I use a mouse tracking paradigm to investigate the online processing of sentences that display a cumulative/distributive ambiguity. Specifically, my goal is to assess whether, as suggested in the semantic literature (see Chapter 1), cumulative readings are obtained automatically ("by default"), whereas distributive readings are derived from more basic interpretations by inserting a distributivity operator.

The approach developed in this chapter relies on similarities between the negative/positive contrast investigated by Dale and Duran (2011) and the cumulative/distributive one. While negative and positive sentences differ on the presence of an *overt* operator that reverses the sentence truth-conditions, cumulative and distributive readings are assumed to differ on the presence of a *covert* distributivity operator. If distributive interpretations involve a "two-step" derivation—the cumulative reading is first derived and the *D* operator is inserted in a second step—, their processing pattern is expected to have an impact in mouse trajectories comparable to the one of negative sentences.

Spelling-out the predictions Following a procedure similar to Dale and Duran's (2011), participants were asked to determine the truth-value of a sentence in a situation depicted graphically. Responses were made by clicking on "true" or "false" buttons, allowing the recording of mouse-movements during each trial.

In critical cases, ambiguous sentences (e.g. 'Two hearts are above two squares') were paired with images that could make the sentence true under a cumulative or under a distributive interpretation, giving rise to *cumulative* and *distributive* conditions. Participants' true responses in each condition are therefore taken to reflect the derivation of the corresponding reading. Conversely, false responses could in principle indicate accessing rates for the alternative reading to the one made true by the picture. Truth values, however, are known to interact with response patterns, giving rise to *congruence effects*, whereby false statements are easier to verify than true ones (Dale & Duran, 2011; Wason & Johnson-Laird, 1972). For this reason, analyses of response times and mouse trajectories are performed only for *true* cumulative and distributive trials.

Under the assumption that cumulative readings are basic and distributive readings derived, we predict an asymmetry between cumulative and distributive conditions. First, since cumulative readings should be easier to compute, response times for true cumulative trials are expected to be lower than for distributive trials. A contrast in mouse trajectories is additionally expected between the two conditions. If distributive readings are derived in a “two-step” manner where basic cumulative readings are initially computed (cf. negative sentences), mouse paths corresponding to true distributive trials should pattern with both negative and *switched* trajectories from Chapter 8. This is expected to contrast with mouse trajectories corresponding to true cumulative trials.

Note that no specific prediction can be made about the accessing rates of each reading (i.e. the percentage of true responses per condition): given that sentences are ambiguous, participants’ preferences might guide them to alternative responses.¹

9.2 Methods

9.2.1 Participants

59 native speakers of English (F= 35) were recruited using Amazon Mechanical Turk. Participants who reported using a device different from a mouse were excluded from the analysis (n=9). All of them were compensated for their participation, which required approximately 20 minutes.

9.2.2 Interface

The interface was the same as the one described in Chapter 8. In each trial, mouse movements were recorded between start- and response-clicks. Pairs of (x, y) -pixel coordinates were saved together with the time.

9.2.3 Procedure

Participants had to evaluate sentence-image pairs and decide, by clicking on the response boxes, whether the image made the statement true or false. They were instructed to do the task as fast as they could and to press the “start” button at the bottom center of the screen in order to start the trial. The procedure is illustrated in Figure 9.1.

The experiment included a short practice phase, where participants received feedback regarding their response and their timing. Specifically, they were asked to start moving faster when they had not initiated any movement within 400ms. This feedback disappeared during the actual experiment. During the experiment, participants received a warning if they had not responded after 7000ms of onset. These trials were not taken into account for the analyses.

9.2.4 Materials

Items consisted on sentence-image pairs. On top of target trials, fillers and controls were created with particular combinations of sentences and pictures.

Sentences Target sentences were constructed automatically using the form in (1):

- (1) Two <shape1> are <predicate> exactly <N> <shape2>.

¹As discussed in Chapters 2 and 4, the fact that cumulative readings are more basic does not necessarily imply that they should be preferred across the board.

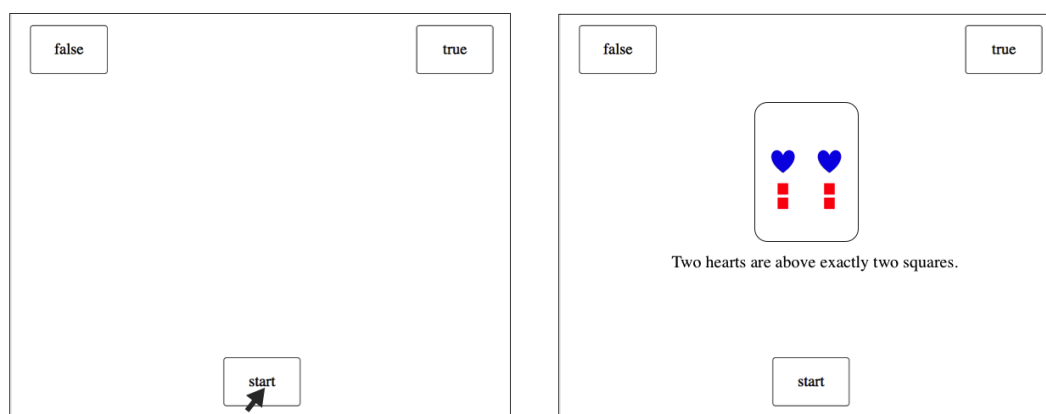


Figure 9.1: Illustration of experimental procedure. After clicking on the “start” button, sentences and images were presented simultaneously and mouse movements started being recorded until response.

Shapes were squares, circles, triangles and hearts, and they were selected randomly in either syntactic position. The numeric term (N) was either ‘one’, ‘two’, ‘four’ or ‘eight’, and its selection was partially determined by the experimental condition (three out of four per condition), as illustrated in Table (9.1) —i.e., not all possible combinations were presented. The predicate was randomly selected among three possibilities: ‘above’, ‘below’ or ‘connected to’. The adverb ‘exactly’ was always included to avoid *at least* interpretations of numeric expressions.²

N TERM IN SENTENCE	N TERM IN PICTURES		
	2	4	8
‘one’	Distributive	–	–
‘two’	Cumulative	Distributive	–
‘four’	–	Cumulative	Distributive
‘eight’	–	–	Cumulative

Table 9.1: Sentence-picture pairs per target condition as a function of the number of shapes in the picture and the numeric quantifiers in the sentence.

Images For each target sentence, we created three types of images: (a) **cumulative**, consistent with the cumulative reading of the sentence; (b) **distributive**, consistent with the distributive interpretation; and (c) **neither**, consistent with no sentence interpretation. An example of the different pictures corresponding to the sentence ‘Two hearts are above exactly two squares’ is presented in Table 9.2.

Control sentences To test the sensitivity of the mouse-tracking method in our specific design, we included control sentences for which effects in mouse-paths are mostly established. We created positive and negative sentences following the frame in (2) (cf. Dale & Duran, 2011). Predicates were randomly selected among two possibilities: ‘at the bottom’ or ‘at the top’.

(2) The <shapes> are (**not**) <predicate>.

These control sentences were paired with the pictures described above, such that the picture could make the sentence either true or false.

²If numerals are interpreted on their “at least” reading, trials in the distributive condition would make cumulative readings of the sentence true as well (i.e. *Two hearts are above at least two squares*).

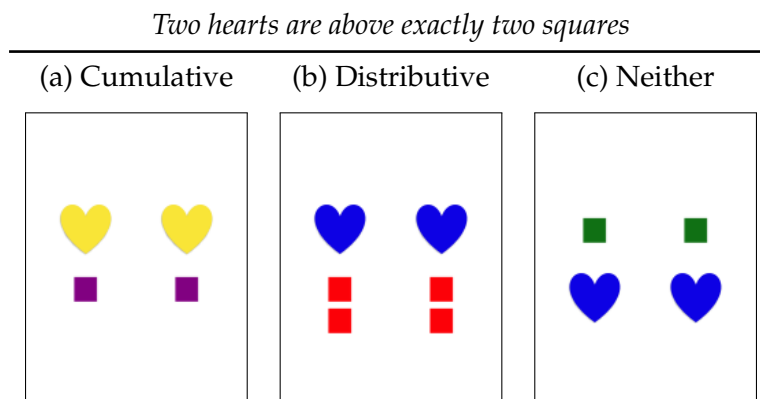


Table 9.2: Example of pictures. Cumulative and distributive pictures only differed on the number of shapes (n term). Neither pictures were created by making false the predicate meaning (e.g. *above* instead of *below*).

9.2.5 Design

Target trials (sentence-picture pairs) expressed one of two possible CONDITIONS (*cumulative* or *distributive*), depending on whether a cumulative or a distributive picture was paired with the target sentence. There were 30 repetitions for each of these conditions, for a total of 60 target trials. To balance the overall amount of true and false trials, there were 60 additional *neither* fillers, where target sentences were paired with neither pictures. We additionally included 120 control trials, which could instantiate one of two SENTENCE TYPES (positive or negative) and one of two TRUTH-VALUES.³

9.3 Results and Discussion

9.3.1 Pre-processing of the data

The time course of mouse trajectories was normalized into 101 time steps, in the same way as in Chapter 8. These normalized trajectories were used as input for the analyses.

Online experiments can give rise to high variability in response times (henceforth, RTs). Even though trials with RTs higher than 7000ms were automatically considered “null responses”, higher values can be found as a result of low temporal resolution in participants’ browser. In order to weight extreme response time values and reduce skewness, response times were log-transformed. After this transformation, the mean and the standard deviation were calculated on these data (7.93, +/- 0.41 s/s, corresponding to approximately 3 seconds). Items were considered outliers if their log-transformed RTs were two SDs outside the mean. Following this constraint, 4% of the trials were excluded.

9.3.2 Analyses

Neither fillers were not analyzed (their accuracy rates were above 85% for all subjects).

Categorical data such as response rates were analyzed by modelling response-type likelihood with logit mixed-effect models (Jaeger, 2008). Response times were fitted into a linear mixed-effects regression model (Baayen, Davidson, & Bates, 2008). In both cases, a maximal random effect structure was kept when possible (Barr et al., 2013). Likelihood ratio tests were used to obtain p -values by comparing more complex models to simpler ones.

³The complete design of the experiment included a third type of trial, which aimed to control for the existence of “at least” cumulative readings. I will not analyse these items here.

In line with the conclusions drawn in Chapter 8, mouse-trajectories were analysed with three different measures: (a) LDA measure, obtained by applying the LDA classifier trained with simulated decisions from the validation experiment; (b) Maximal Deviation (MD), obtained by extracting the point in each trial that maximizes the distance between the path and the ideal trajectory; and (c) Maximal Log-Ratio, obtained by extracting the point that maximizes the ratio between the distance to the the target and to the alternative response.

To assess how different the two experimental conditions are, I followed the procedure described in Chapter 8. First, the performance of each measure was evaluated by calculating the area under the ROC curve (henceforth, AUC) for 1000 new samples bootstrapped from the original data set. These AUC values are an indication of how well the measure distinguishes between the two experimental conditions. As a second step, the AUC values were compared to the ones one would have obtained from applying the same measure to a set of data where there is no difference between experimental conditions —i.e. a ‘permutation distribution’ obtained by shuffling the experimental labels. The separability between these two distributions of AUC values (observed distribution vs. permutation distribution) is taken as an estimation of whether or not there is a true difference between experimental conditions.

Before summarising the results, let me note that the total amount of trials in this experiment is 20 times larger than the one in Dale and Duran (2011) (200 trials and 16 trials, respectively). Given that the impact of experimental manipulations in online measures such as RTs and mouse trajectories is susceptible to either decrease or disappear after repetition, all the analyses were applied to the complete set of data as well as to a subset containing only the first 20 control/target trials. This restriction, however, did not make any substantial difference. I will thus report the results corresponding to the whole data set.

9.3.3 Control results

Mean accuracy rates were above 95% for both *positive* (98%; SE=.3) and *negative* (95%; SE=.7) control trials. The averaged paths for accurate controls in the four conditions are given in Figure 9.2. These trajectories suggest a difference between *true* positive and negative trials, whereby paths corresponding to negative trials are less straightforward than their positive counterparts. This difference, however, is visually smaller than the one obtained in Chapter 8 (see Figure 8.11).

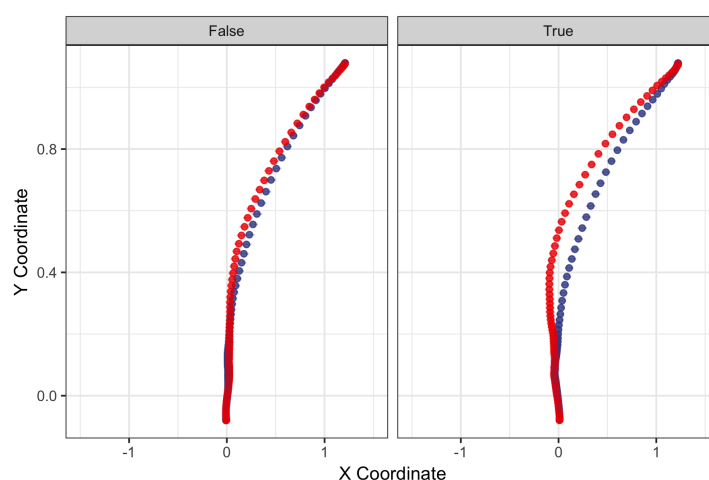


Figure 9.2: Controls. Mean trajectories of **positive** and **negative** accurate trials.

Following the procedure in Chapter 8, the analyses were restricted to *true* positive and negative controls. The LDA classifier trained with simulated decisions from the validation experiment was applied to this data, and the Maximal Deviation and LogRatio points were

computed for each trial. Figure 9.3 shows the distribution and means for the resulting values.

As observed in Section 8.3.1, the LDA measure obtained for each trial corresponds to the trial position in the LDA classification axis. In this axis, zero represents the midpoint between the two classes: when a path underlies a *straightforward* decision it will be assigned a negative LDA value, whereas when it underlies a *switched* decision, the LDA measure will be positive. The distribution of the LDA measure here suggests that most trajectories for both experimental conditions underlie straightforward decisions. Still, there might be more decision changes underlying negative than positive trials.

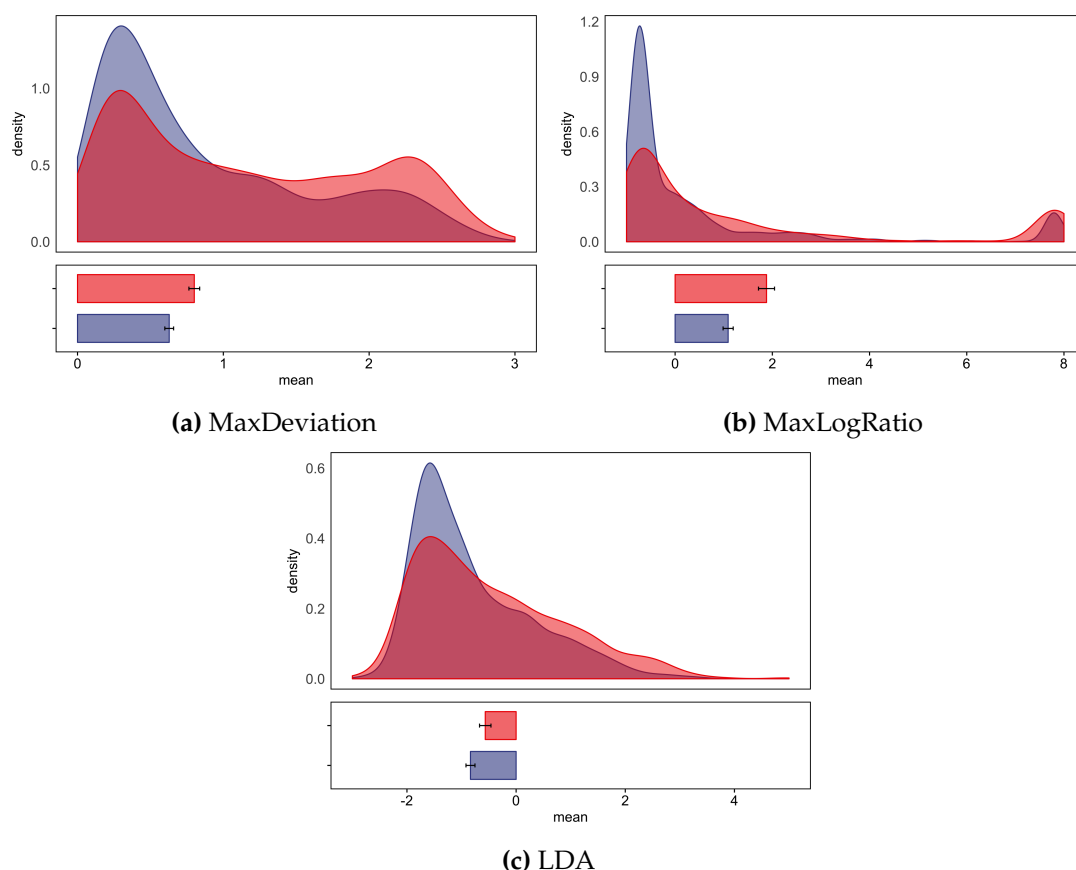


Figure 9.3: Controls. LDA measure, Maximal Deviation and Maximal LogRatio for *true positive* and *negative* trials. Error bars correspond to the mean standard error.

To assess how well these three measures distinguish between positive and negative trials, we calculated the AUC for 1000 bootstrapped samples from the original data. The distribution of these AUC values is shown in Figure 9.4 (red curve). Mean AUC values are indicated by the horizontal red line (LDA = .56; Maximal Deviation = .59; Maximal LogRatio = .6). Note that these mean values are generally lower than the ones obtained in Dale and Duran's (2011) replication. This is consistent with the visual pattern observed in mean trajectories (Figure 9.2), which suggests a smaller difference between the two conditions.

The distribution of the observed AUC values was compared to a 'permutation distribution' (grey curve in Figure 9.4), which is the distribution of AUC values obtained from applying each measure to a set of data where the experimental labels (positive, negative) were scrambled. For the three measures, the separability between the two distributions is complete, as indicated by a new calculation of the AUC, now on these two distributions (LDA = .99, MD = 1, MaxLogRatio = 1).

These findings suggest that there is a real difference between mouse trajectories corresponding to positive and negative trials, such that the verification of negative sentences gives rise to more deviated paths than the verification of their positive counterparts. Specifically, the LDA

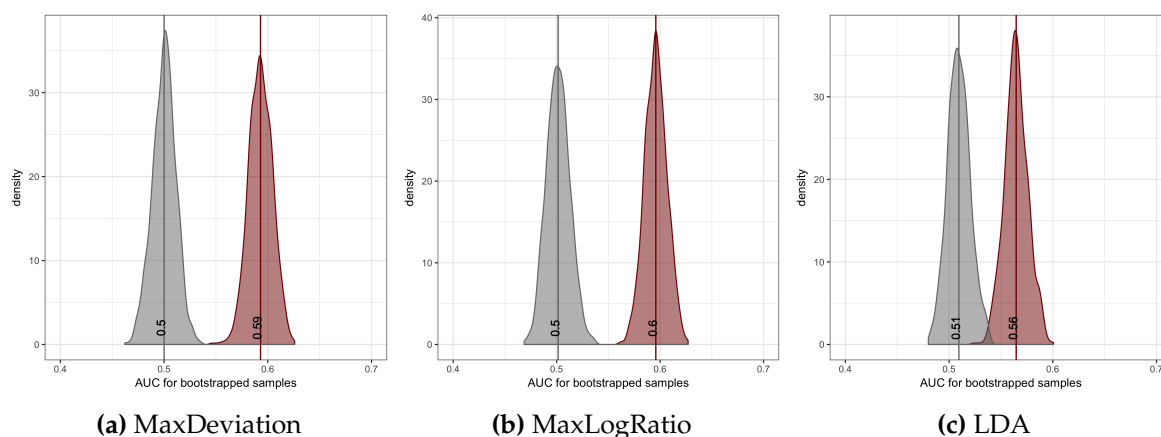


Figure 9.4: Controls. Distribution of AUC values for **observed data** versus **scrambled data**.

results indicate that, while most trials in both conditions involve straight decisions, more decision changes are generated by negative than by positive sentences. This difference is captured by the three measures (LDA, MD, MaxLogRatio), but to a different extent: the distinction made by the LDA seems to be weaker than the one made by the Maximal Deviation and LogRatio, as suggested by the mean AUC values.

As observed, the difference between conditions is itself smaller than the one obtained in Chapter 8. This fact might be related with the experimental set-up. In particular, there are two main differences between Dale and Duran's (2011) experiment and the present study which might be responsible for the reduction of the effect. First, while in Dale and Duran's experiment participants had to judge simple statements based on their world knowledge, in the present study they had to verify the truth of a sentence with respect to a picture. These two tasks might involve very different cognitive processes. For instance, a sentence-picture verification task might be performed by first translating the content of the sentence and of the picture into the same representational format, to then be able to compare them (H. H. Clark & Chase, 1972). Instead, judging statements against world knowledge might demand retrieving and generalizing memorised information. As a result, these two tasks might have associated different patterns of response times or mouse trajectories.

Moreover, the present study was 20 times longer than the one by Dale and Duran and it included additional sentence types (e.g. target trials). Even if restricting the analyses to a subset of the data did not make a difference in the results, it is unclear how the rest of the trials might have influenced participants' performance on controls.

9.3.4 Target results

Accessing rates Figure 9.5a illustrates the mean proportion of *true* responses per target condition. Participants' true responses are significantly higher for cumulative than for distributive trials ($\chi^2 = 89, p < .001$).⁴ Specifically, true responses for cumulative trials are at ceiling ($\sim 94\%$), suggesting that cumulative readings are universally accepted. Instead, distributive readings seem to be accessed only marginally ($\sim 27\%$ of true responses).

The marginal status of distributive interpretations is somehow surprising, as these readings have been shown to be fully available for similar sentences in other experiments (cf. Chapters 2, 3 and 7). This pattern of results, however, is consistent with the idea that the probability of assigning a given interpretation to an ambiguous sentence is not stable, but depends on a range

⁴An additional model indicated a significant interaction Condition \times Predicate ($\chi^2 = 43; p < .001$) in response rates. In particular, the rate of true responses in distributive trials increases for *connected* to predicates in contrast with both *above* and *below* configurations ($\chi^2 = 23, p < .001; \chi^2 = 23, p < .001$).

of factors, varying drastically across experiments (see discussion in Part I). It is conceivable, for instance, that the fact that participants were instructed to perform the experiment as fast as they could (under time pressure) made them develop a strategy consisting on deriving always the easiest interpretation, which in this case would be the cumulative reading.

Alternatively, the large difference in accessing rates between experimental conditions might raise the question of whether true responses to distributive trials do indeed reflect the actual rate of distributive readings or whether they should rather be taken as errors. In other words, can these findings be interpreted by saying that participants are *always* accessing cumulative interpretations, and making some mistakes on the way? This does not seem to be the case. To the extent that simple errors occur independently of the condition, the same percentage of errors should be attested in both cumulative and distributive conditions. The percentage of true responses in the distributive condition should then be comparable to the the proportion of ‘false’ responses for cumulative trials, contrary to fact. We can therefore conclude that participants are marginally deriving distributive interpretations.⁵

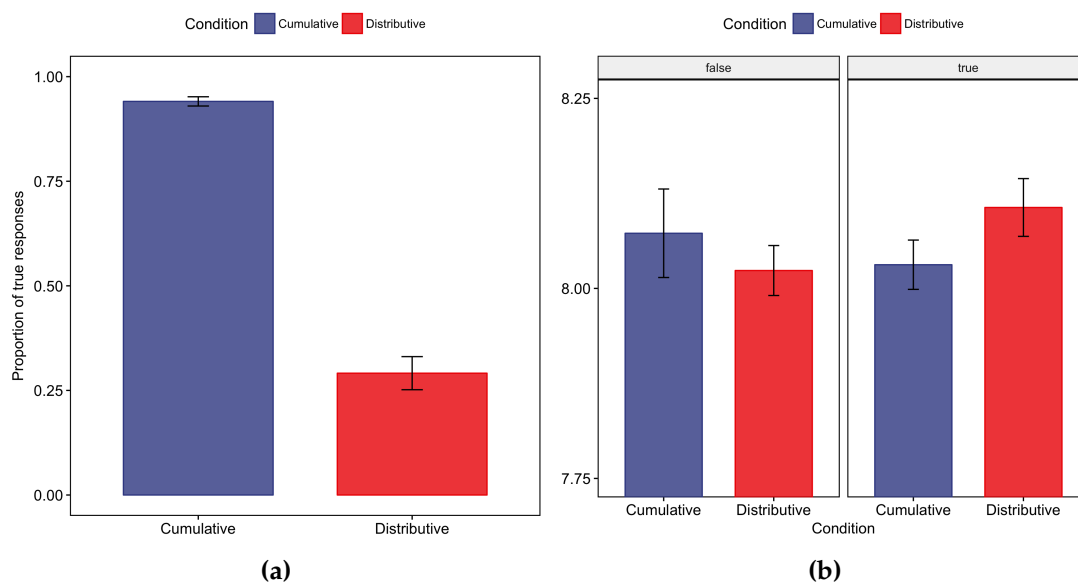


Figure 9.5: Targets. (a) Acceptance rates quantified as proportion of true responses per condition. (b) Log-transformed Response Times per condition and response. Error bars correspond to the mean standard error.

Response times Response times were analyzed in log-units only for true trials (Figure 9.5b). Participants were significantly faster in cumulative than in distributive targets (main effect of Condition: $\chi^2 = 10.5; p = .001$). The fact that distributive readings are only marginally available makes this difference in RTs difficult to interpret. Response latencies are known to correlate with processing cost, suggesting that saying ‘true’ is cognitively more effortful in distributive than in cumulative trials. It is, however, unclear whether the difficulty associated with true distributive trials results from the specific derivation of distributive readings, or just from accepting a drastically dispreferred or infrequent interpretation. In other words, a slow-down in responses could be related either with a more demanding semantic computation, which may involve the insertion of an additional operator, or with difficulty to accept that there is a reading under which the sentence could be accepted.

⁵Note that these results support the view that *false* responses to distributive trials correspond to accessing rates for cumulative readings. Due to the aforementioned congruence effect, I will not take these trials into account for the analyses on response times and mouse trajectories.

Mouse tracking data Averaged mouse-trajectories per condition are shown in Figure 9.6.

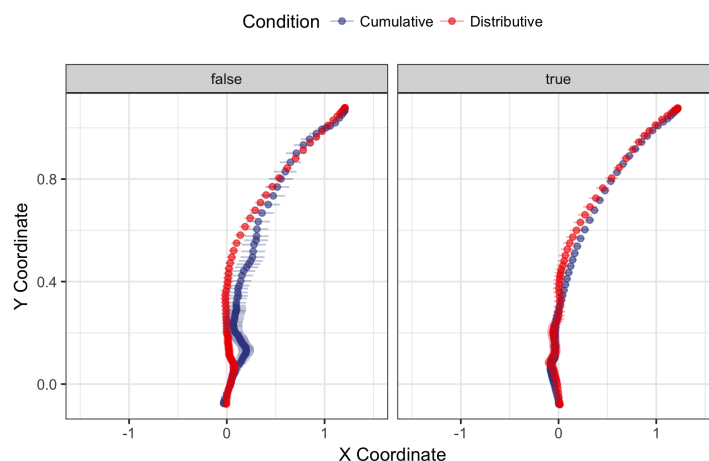


Figure 9.6: Targets. Mean trajectories for **cumulative** and **distributive** trials.

The findings obtained for control trials suggest that the three measures proposed to analyze mouse trajectories —LDA measure, Maximal Deviation and Maximal LogRatio— can capture differences between experimental manipulations in this specific experimental set-up. The distribution and means obtained for true distributive and cumulative trials after applying the LDA classifier, and computing the Maximal Deviation and LogRatio are shown in Figure 9.7.

Are these measures capable of separating the two experimental conditions? This question was addressed by applying the same procedure as for control trials: for each measure, we calculated AUC values for bootstrapped samples from the original data set, and contrasted these values against a null hypothesis of no difference between experimental conditions (i.e. permutation distribution, see Figure 9.8). Mean AUC values for the observed data indicate that the three measures make a worse distinction between cumulative and distributive targets than between negative and positive controls (LDA=.52; Maximal Deviation=.52; Maximal LogRatio=.52). Moreover, the observed AUC values are not different from the ones obtained by shuffling the experimental labels: the separability between the two distributions is nearly at chance (LDA=.58; Maximal Deviation=.59; Maximal LogRatio=.68). This reveals that the three measures fail to distinguish between mouse trajectories from true cumulative and distributive trials.

The absence of evidence of a difference between conditions should not be taken as indicating that the derivation of distributive and cumulative readings has the same processing pattern —i.e. it is not evidence of absence of a difference. Likewise, these null results do not necessarily imply that the online processing of plural ambiguous sentences cannot be studied by means of a mouse tracking method. Instead, there are a number of properties of this specific study that might have influenced these mouse-tracking results. As mentioned for control trials, the use of a sentence-picture verification task and the length of the experiment might have resulted in a weakening of the impact of decision processes into mouse trajectories. Moreover, given that only *true* cumulative and distributive trials were taken into account for the analyses, there was an imbalanced proportion of trials for each condition (due to the marginality of distributive readings). The influence that this (dis)-proportion might have in the analyses of mouse paths is difficult to estimate.

Before concluding, let me note as an interesting finding that mouse trajectories do not mimic response times. In particular, mouse paths do not provide more information than response times, despite the fact that they actually contain time information. Indeed, a difference between true distributive and cumulative trials was observed in response latencies. As discussed, we cannot identify at the moment what might be the source of this difference —semantic

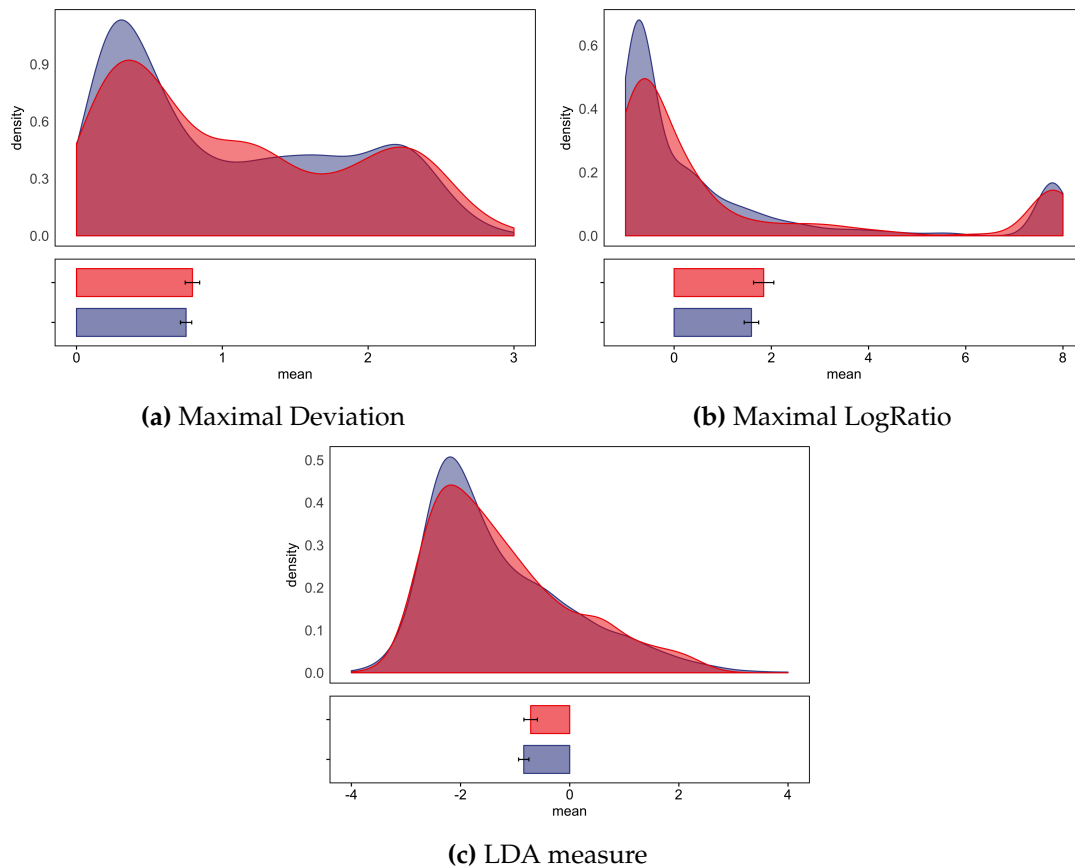


Figure 9.7: Targets. LDA measure, Maximal Deviation and Maximal LogRatio for *true* **cumulative** and **distributive** trials. Error bars correspond to the mean standard error.

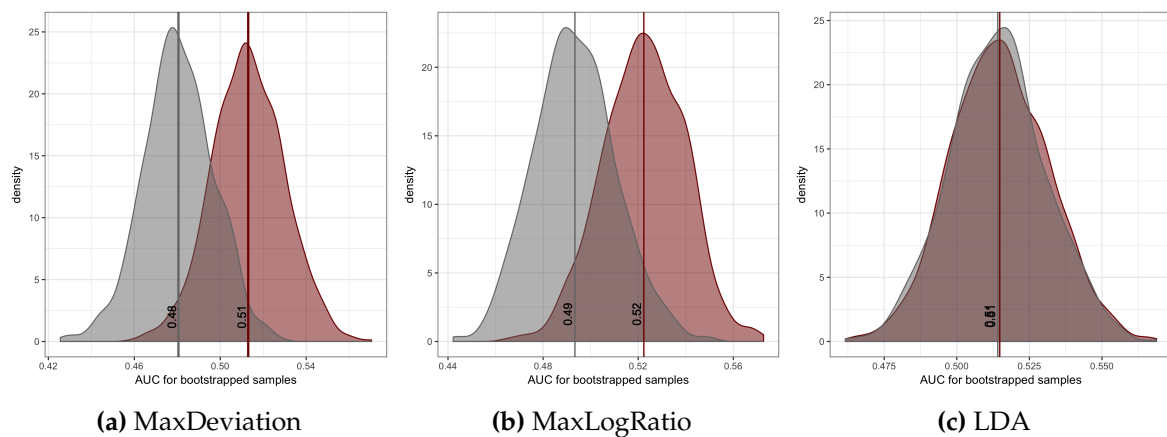


Figure 9.8: Targets. Distribution of AUC values for **observed data** versus **scrambled data**

processing or reflection of preference patterns. In any case, our mouse-tracking results reveal that mouse-trajectories are not sensitive to this difference.

9.4 Conclusions

The main goal of this chapter was to investigate how different interpretations of plural ambiguous sentences are derived during online comprehension. To address this question, the tools developed in Chapter 8 were applied to mouse-tracking data obtained from a sentence-picture

verification task involving plural ambiguities. While the findings reported here do not allow us to draw specific conclusions about the online derivation of cumulative and distributive interpretations, they do provide some methodological insights regarding the use of a mouse-tracking method in sentence-verification tasks.

First, the impact of experimental manipulations in mouse trajectories seems to be quite sensitive to the experimental set-up. Verifying a sentence against a picture, for example, might involve some additional processing cost, leading participants to wait until all the information has been integrated before making the decision. As a result, mouse paths might not be able to capture the dynamics of online processing. Presumably, simpler experimental set-ups might allow participants to execute their responses in parallel to semantic processing.

Moreover, in order to explore the online derivation of alternative interpretations of ambiguous sentences, one needs to ensure that the relevant readings are fully available to participants. This is independent of whether alternative interpretations have different preference patterns. In other words, the processing pattern associated to, for example, distributive interpretations cannot be investigated if distributive readings are never derived in the first place.

Finally, as suggested in Chapter 8, the LDA classifier (together with more traditional measures such as the Maximal Deviation and Maximal LogRatio) has shown to be sensitive to contrasts between experimental conditions, even when differences in mouse trajectories are small. Note, however, that the more complex is the experimental task, the worse is the performance of the LDA classifier. In principle, this drop in performance might be prevented by training the classifier with a broader set of data, which should include mouse paths underlying not only more complex decision processes but also experimental settings more similar to the ones used in the test set.

Part IV

Coming back to plurals: The semantic import of morphological number

I come back to issues related to plurality, and provide a theoretical contribution to the debate about the denotation of plural morphology. I present a study on the semantic import of plural marking in Spanish interrogative words. My proposal here reveals that current approaches to number marking need to be refined in order to account for cross-linguistic and within-language variation.

10 | Strong plurality and d-linking in Spanish interrogatives*

Maldonado, M. Strong plurality and d-linking in Spanish interrogatives. Submitted.

Abstract

What is the semantic import of number morphology? This question has been traditionally addressed by focusing on singular and plural noun phrases. The present work brings interrogative phrases into the picture. We analyse Spanish bare interrogative ‘quién’ and its plural counterpart ‘quiénes’. Unlike *which*-questions in both English and Spanish, the behaviour of *quién*- and *quiénes*-interrogatives cannot be easily explained by most accounts of semantic number. In contrast, we argue that the distribution of these interrogatives in Spanish can be well accounted for by assuming that the plural ‘quiénes’ triggers a *strong plurality presupposition*, and can only be used in *d-linking* contexts, whereas ‘quién’ carries no specific requirement, as far as its semantics is concerned. As a result, our proposal shows that current approaches to number marking need to be refined in order to account for cross-linguistic and within-language variation.

10.1 Introduction

The question of the semantic import of overt number morphology is at issue in current semantics research (de Swart & Farkas, 2010; Link, 1983; Sauerland, 2003; Spector, 2007a; Zweig, 2009, a.o.). While most approaches have addressed this question by discussing number differences in declarative sentences, this paper analyzes the semantics of singular and plural *wh*-interrogatives, which appear to be problematic for most accounts of both plural and question semantics.

Singular and plural *which*-interrogatives in English differ in what the speaker can expect to be a *complete true answer* to her question. A similar pattern is attested for Spanish *cuál*-interrogatives.¹

- | | | | |
|-----|----|---|-------------------------|
| (1) | a. | Which client called? | ✓ John; ✗ John and Mary |
| | b. | Which clients called? | ✗ John; ✓ John and Mary |
| (2) | a. | Cuál/Qué cliente llamó?
which.SG/what client called?
'Which client called?' | ✓ John; ✗ John and Mary |

For helpful suggestions and comments, I thank Amir Anvari, Emmanuel Chemla, Milica Denic, Manuel Kriz, Irene Heim, Roger Schwarzschild, Andrés Soria Ruiz, and the audience of Console 25 at Leipzig University. Special thanks to Benjamin Spector for discussions during the whole development of this work. This work was supported by ANR-11-0001-02 PSL, ANR-10-LABX-0087 IEC and ANR-14-CE30-0010- 01 TriLogMean. Of course, all errors are my own.

¹We will treat interrogatives headed by ‘cuál’ (‘which’) and by ‘qué’ (‘what’) in Spanish as equivalent to each other. As in English, they do contrast in their d-linking requirements.

- b. Cuáles/Qué clientes llamaron?
 which.PL/what clients called?
 'Which clients called?' ✗ John; ✓ John and Mary

Singular *wh*-questions trigger an *uniqueness presupposition*: questions such as (1a) and (2a) can only be uttered if the speaker believes exactly one client called (henceforth *exactly-one* contexts; Dayal, 1996). Conversely, uttering a plural interrogative such as (1b) or (2b) triggers the inference that the speaker believes that more than one client called (*plurality inference*). These different inferences can be attributed to differences in number morphology.

Interestingly, English *who*-interrogatives such as (3) do not trigger any specific inference about how many clients the speaker believes to have called: they have associated neither a *uniqueness presupposition* nor a *plurality inference*. Despite its singular morphology, these questions do not pattern with singular or with plural *which*-interrogatives.²

- (3) Who called? ✓ John; ✓ John and Mary

The puzzling behaviour of *who*-interrogatives might be related to the fact that the quantifier 'who' lacks a plural counterpart in English (Chierchia, 1993; Hagstrom, 2003): 'who' can then be thought of as semantically underspecified for number, ranging over both atomic individuals and pluralities. The fact that 'who' triggers singular agreement with the verb could then be understood as default agreement.

Although the implementation of this idea might differ across accounts (see following section), one can easily notice that a specific prediction is made about languages that distinguish between singular and plural 'who': in these languages, the contrast attested for singular and plural *which*-interrogatives is expected to also arise for *who*-interrogatives. Specifically, singular *who*-interrogatives should trigger a uniqueness presupposition, whereas plural *who*-interrogatives should trigger a plurality inference. Spanish is one of such languages, making a morphological distinction between *who*_{SG} ('quién') and *who*_{PL} ('quiénes'):

- (4) a. Quién llamó?
 who.SG called.SG?
 'Who called?' ✓ John; ✓ John and Mary
- b. Quiénes llamaron?
 who.PL called.PL?
 'Who called?' ✗ John; ✓ John and Mary

As illustrated in (4a), singular *quién*-questions are compatible with both plural and singular answers. In other words, they appear to behave just like *who*-questions in English, suggesting that 'quién' is not semantically singular (no uniqueness presupposition). In contrast, the plural alternative in (4b) does trigger a plurality inference such that the speaker believes that more than one person called. This inference makes the question incompatible with singular answers.

The Spanish data in (4) presents a challenge for a unified semantics of number morphology: even in presence of a plural alternative, singular morphology does not have the same semantic import across different *wh*-phrases, as revealed by the contrast between (1a)/(2a) and (4a). As we will see in the following section, most semantic accounts of number morphology consider singular and plural morphemes as alternatives to each other. Within these frameworks, having an unstable meaning for singular marking will cause unstable semantics for plural morphology.

The purpose of this paper is to provide an explanation for the distribution of *quién* and *quiénes*-interrogatives in Spanish. We will argue that the pattern illustrated in (4) can be derived directly from requirements of the plural 'quiénes'. In particular, our account builds on two main

²For instance, 'who' can be combined with collective predicates (e.g. *Who gathered in the hallway?*), indicating that this interrogative quantifier is not semantically singular; namely it can range over both atoms and pluralities.

assumptions: (a) a strong semantics for the plural ‘*quiénes*’, which triggers a *strong plurality* presupposition and can only be used in *d-linking* contexts; and (b) a weak semantics for the singular ‘*quién*’, which carries no specific requirement, and can be used whenever its plural alternative is not available. By relying on a strong account of plurality, our proposal challenges current theories of *weak* plural meaning and their cross-linguistic generality.

This article is organized as follows: We will start by presenting some background assumptions about number marking in interrogative sentences (Section 10.2). Then, Section 10.3 puts forward our characterization of Spanish *quién* and *quiénes*-interrogatives, which is based on cardinality (Section 10.3.1) and d-linking (Section 10.3.2) requirements. The complete account for the data is provided in Section 10.4. Finally, Section 10.5 concludes and comments on open questions for further research.

10.2 Number marking in interrogative sentences

The English contrast illustrated in (1) can be well-accounted for by making two assumptions: (A) plural marking has weak semantic import (i.e. *weak account of plurality*, Link, 1983; Sauerland, 2003; Spector, 2007a); and (B) interrogatives presuppose the existence of a maximally informative true answer (Dayal, 1996).

Current approaches to plurality stipulate, following Link (1983), that the domain of individuals D_e is closed under sum (i.e. it contains atomic individuals as well as their sums). However, they do differ on how exactly plural and singular predicates draw their reference from the domain. *Weak accounts of plurality* (cf. A) assume that the meaning of plural morphology is strictly weaker than the one of singular: while singular predicates (e.g. ‘client’) range over atomic individuals, plural predicates (e.g. ‘clients’) range over both atoms and pluralities. Accounts might differ in how they derive this weak plural meaning (see Landman, 2000; Sauerland, 2003; Sauerland et al., 2005; Spector, 2007a; Zweig, 2009 for different proposals). For the sake of simplicity, here we will follow the view proposed in Sauerland (2003). Note, however, that alternative implementations will not make any substantial difference to our proposal. On Sauerland’s view, the singular morpheme presupposes that the extension of the predicate to which it is applied only contains atomic individuals. In contrast, the plural morpheme has no semantic contribution (i.e. it is semantically vacuous).³

Assuming that the denotation of a question is the set of propositions that count as its possible answers (Hamblin, 1973), a singular question such as (1a) will only include in its denotation singular propositions of the form ‘ x called’, with x being an atomic individual. This is a direct consequence of the singular morphology in the predicate ‘client’. Plural questions will range over both singular and plural propositions. An illustration of the denotation of (1a) and (1b) is provided in (5). Following Heim and von Stechow’s (2016) reconstruction of Karttunen’s (1977) compositional semantics, we assume *wh*-words to be existential quantifiers and derive the question denotation by posting an abstract complementizer that acts as a “proto-question” operator. For simplicity, the requirements of the singular morpheme are assumed to be inherited by the proposition.

$$(5) \quad a. \quad \llbracket (1a) \rrbracket^w = \lambda p. \exists x. \text{atom}(x) \wedge \text{client}(w)(x) \wedge p = \lambda w'. x \text{ called in } w' = \left\{ \begin{array}{l} \lambda w'. m \text{ called in } w', \\ \lambda w'. b \text{ called in } w', \\ \lambda w'. j \text{ called in } w' \end{array} \right\}$$

³Sauerland furthermore assumes the existence of a star-operator (*) that closes the predicate under sum formation (in the lines of Link, 1983): if the predicate is true of $\{a, b\}$, the star operator would make the predicate true of $\{a, b, a \oplus b\}$. This operator is compatible with the plural morpheme and incompatible with the singular.

$$b. \quad \llbracket (1b) \rrbracket^w = \lambda p. \exists x. \text{clients}(w)(x) \wedge p = \lambda w'. x \text{ called in } w' = \left\{ \begin{array}{l} \lambda w'. m \text{ called in } w', \\ \lambda w'. b \text{ called in } w', \\ \lambda w'. j \text{ called in } w', \\ \lambda w'. j \oplus m \text{ called in } w', \\ \lambda w'. j \oplus b \text{ called in } w', \\ \lambda w'. b \oplus m \text{ called in } w', \\ \lambda w'. j \oplus m \oplus b \text{ called in } w' \end{array} \right\}$$

By the assumption in (B), the set of *true* answers to a question should contain a member that entails all the other members. This member will be the *maximally informative* true answer to the question. Following Dayal (1996), the requirement in (B) can be captured in terms of the answerhood operator defined in (6): a question Q is felicitous only if in all worlds w compatible with common knowledge, $ANS(Q)(w)$ is defined.

- (6) $ANS(Q)(w) = \iota p [p \in Q \wedge p(w) \wedge \forall p' \in Q [p'(w) \rightarrow p \subset p']]$
 In words: $ANS(Q)(w)$ is the maximally informative answer to Q in w , if there is one, and is undefined otherwise.

Since singular questions can only have such a maximally informative answer if the set of true propositions is a singleton (' x called' does not entail ' y called' unless $x = y$), the *uniqueness presupposition* is directly captured by the presuppositions of the ANS operator. In the case of plural, the set of true propositions can contain more than one member (' $x \oplus y$ called' entails ' x called' and ' y called'). However, once we assume that plural questions have in their denotation both plural and singular propositions (cf. (5b)), nothing prevent us from uttering a plural interrogative in a context where a (maximally informative) *singular* answer is expected. This is not a welcome result since, as we have observed, plural questions tend to require a plurality named in their answer (i.e. the aforementioned *plurality inference*).

By analogy with the case of nominal expressions, the plurality inference associated with plural interrogatives can be thought to arise as an *implicated presupposition* (Heim, 1991; Sauerland, 2008), product of the competition between singular and plural alternatives. The basic reasoning goes as follows: In every scenario where two sentences are contextually equivalent (7a), a pragmatic principle such as *Maximize presupposition!* can apply, requiring to use of the item with stronger presuppositions.

- (7) a. Two sentences F and F' are *contextually equivalent* in context c iff in every world w of c , $F(w)$ and $F'(w)$ are defined and $F(w)$ and $F'(w)$ have the same truth value.
 b. MAXIMIZE PRESUPPOSITION!
 If F and F' are contextually equivalent in c and the presuppositions of F are stronger than those of F' , then one must use F and not F' .

The notion of contextual equivalence has to be slightly modified to be applied to interrogative sentences: two interrogatives will be *contextually equivalent* in one context if they have the same maximally informative answer in every world compatible with the context; namely, if ANS returns the same exhaustive true answer.⁴ For concreteness, we will then consider ANS to be incorporated into the LF of matrix questions. Whenever it is presupposed that exactly one person went to the party, the two interrogatives in (1) are contextually equivalent, and the speaker should use the alternative carrying stronger presuppositions; namely, the singular.

Given that using the plural interrogative is only appropriate in contexts where it is not common knowledge that there is a unique singular answer, these questions trigger the inference that the uniqueness presupposition is not satisfied: (1b) is then felicitous whenever its singular counterpart (1a) is not. As a result, plural interrogatives such as (1b) are predicted to have the implicated presupposition that it is *not* common knowledge that exactly one person called. That

⁴Two interrogative sentences F and F' are *contextually equivalent* in context c iff in every world w of c where $ANS(F)(w)$ and $ANS(F')(w)$ are defined, $ANS(F)(w)$ and $ANS(F')(w)$ return the same maximally informative answer.

is to say, *it is possible*, according to common knowledge, that more than one person have called.

In the same way as with implicated presuppositions for declarative sentences, whenever the speaker is considered to be reliable and knowledgeable, the logic of the epistemic-step for presuppositions (or anti-presuppositions) developed in Chemla (2008) can be applied, and the inference can be strengthened into a *strong* implicated presupposition, such as “it’s common knowledge that more than one person called” (i.e. it is common knowledge that it is not the case exactly one person called).⁵ Strikingly, the plurality inference for plural *which*-interrogatives is typically an implicated presupposition of this second class (a strong one): when the speaker utters the question, it seems to imply that the complete answer is a plural answer.

It is worth noticing that the distribution of singular and plural forms under a weak account of plurality crucially depends on the asymmetry in presuppositional strength between the two items: the plurality inference for the plural arises as long as the singular has stronger presuppositions; otherwise, a principle such as *Maximize Presupposition!* could not be applied.

To conclude this section, we would like to briefly explore how a weak account of plurality could deal with English *who*-interrogatives. The absence of a uniqueness presupposition for these questions indicates that, despite triggering singular agreement, the *wh*-element ‘who’ is semantically plural or underspecified (see Dayal, 1996). ‘Who’ should then range over the whole domain of people, and interrogatives such as (3) should denote a set containing both plural and singular propositions, just like the plural question in (5b). The main difference between plural and ‘neutral’ interrogatives would be that only the former has an alternative with stronger presuppositions, and, as a result, no plurality inference is predicted for *who*-interrogatives. These assumptions are hard to integrate with the semantics proposed by Sauerland for number morphology: while ‘who’ phrases appear to receive singular marking, the singular morpheme seems to have the same semantic import as the plural one (i.e. it is vacuous).

10.3 Characterizing *quién* and *quiénes* interrogatives

10.3.1 Cardinality requirement

The distribution of *quién* and *quiénes* interrogatives in Spanish is restricted by conditions on contexts: the possibility of using each of these interrogatives depends on the beliefs the speaker has about the complete answer to her question. The data in (8)- (9) illustrate, together with (4), such distribution.

- (8) *The speaker is pointing towards her clients, who are on the other side of the room.*
- a. Uno de ellos llamó ayer, pero no estoy segura (de) quién (#quiénes).
One of them called yesterday, but not am sure (of) who.SG (#who.PL)
‘One of them called yesterday, but I am not sure who’
 - b. Al menos uno de ellos llamó ayer, pero no estoy segura (de) quién
At least one of them called yesterday, but not am sure (of) who.SG
(#quiénes).
(#who.PL)
‘At least one of them called yesterday, but I am not sure who.’
 - c. Varios de ellos llamaron ayer, pero no estoy segura (de) quiénes
Several of them called yesterday, but not am sure (of) who.PL
(#quién).
(#who.SG)

⁵In the case of questions, there is of course an inference that the speaker does not know the answer to the question. So for the reliability assumption to hold, it has to be the case the hearer believes that the speaker, at the same time, knows whether one or more than one people called, but does not know who exactly called. It is not clear that this line of reasoning is sufficient to explain why, out of the blue, a plural *wh*-question suggests that the answer involves a true plurality, but we will not pursue this question here.

‘Several of them called yesterday, but I am not sure who.’

- (9) Mucha gente de la que invitamos llamó ayer, pero no estoy segura (de)
 Many people of the that invited.1PL called yesterday, but not am sure (of)
 quiénes (?? quién).
 who.PL (?? who.SG)
 ‘Many of the people we invited called yesterday, but I am not sure who’

Singular *quién*-interrogatives such as (4a) can be felicitously uttered in two alternative scenarios: whenever the speaker expects a singular answer (i.e. *exactly-one* contexts in (8a)) and whenever the speaker does not know whether one or more than one person called —e.g. (8b). We will call these contexts *ignorance scenarios* because the speaker is ignorant about the cardinality of the expected answer. The fact that *quién*-interrogatives are available in ignorance scenarios is what makes them fully compatible with both plural and singular answers, as observed in (4).

The plural alternative with ‘*quiénes*’ is not available in either of these two contexts. *Quiénes*-interrogatives can only be felicitous as long as it is common knowledge that a plurality will be named in the complete answer (cf. plurality inference). Specifically, the common ground should contain the proposition “more than one person VP” to admit a interrogative of the form *quiénes VP?*.⁶ In (8c), the speaker asserts that more than one person called, introducing the proposition in the common ground, and questions their identity. *Quién*-interrogatives are degraded in this context. As illustrated in (9), this contrast is partly independent of the morphological marking in the noun: even though the noun ‘*gente*’ triggers singular agreement in Spanish, it is semantically plural and therefore it seems to force the use of the plural ‘*quiénes*’.⁷

The availability of *quién*-interrogatives in ignorance contexts (e.g., (8b)) is well accounted for by an underspecified entry for ‘*quién*’, similar to the one required for ‘*who*’ in English.

- (10) $\llbracket \text{QUIÉN}^{[+WH]} \rrbracket^w = \lambda F_{et}. \exists x. x \in \llbracket \text{people} \rrbracket^w \wedge F(w)(x)$, where $\llbracket \text{people} \rrbracket^w$ is closed under sum.

Since ‘*quién*’ will range over the set of people—which we could consider equivalent to the domain of singular and plural individuals—, *quién*-interrogatives will include in their denotation both singular and plural propositions. No constraint would then be imposed to the context of utterance. For instance, in a scenario where the speaker is ignorant about answer cardinality, the *ANS* operator would be able to return both singular or plural propositions, depending on the world.

The question then is how to derive the plurality inference attested for *quiénes*-interrogatives. Depending on the specific account of number marking one adopts, such inference can be derived in one of two ways: as an implicated presupposition, product of the competition with a stronger alternative (cf. weak accounts of plurality described in Section 10.2), or as part of the literal meaning of the interrogative (i.e. *strong* accounts of plurality).

In their simplest form, weak accounts of plurality predict singular *quién*-interrogatives to give rise to a uniqueness presupposition. The plurality inference attested for the plural should arise as an implicated presupposition. As observed, this first prediction is not borne out by the data: an underspecified entry for ‘*quién*’ is indeed required; therefore, there is at least one entry

⁶We assume the notion of common ground proposed by Stalnaker (1978). A proposition is in the common ground if each interlocutor is “disposed to act as if he assumes or believes that the proposition is true, and as if he assumes or believes that his audience assumes or believes that it is true as well”.

⁷Some speakers report unclear judgments regarding the unavailability of ‘*quién*’ in these last cases, in particular in (9). While they all agree in that the alternative with ‘*quiénes*’ is preferred, they do not necessarily consider the sentence unacceptable. As we will note in Section 10.3.2, the use of ‘*quiénes*’ is constrained by both a plurality and a *d-linking* requirement. The fact that some speakers might find ‘*quién*’ acceptable in these scenarios is presumably grounded on a non-*d-linked* interpretation.

for ‘quién’ which is not capable of competing with a weak plural alternative.

Nonetheless, one might still conceive the possibility that ‘quién’ is actually ambiguous between a form that ranges only over atoms (‘quién’_{SG}), and one that is neutral to semantic number (‘quién’_∅). This ambiguity might emerge from the singular morpheme or be lexically based. In either case, the neutral form ‘quién’_∅ would be the one used in ignorance contexts (8b), whereas ‘quién’_{SG}, equivalent to ‘which person’ in English, would be used in *exactly-one* scenarios (8a). The plural ‘quiénes’ would have a weak semantics, equivalent to ‘which people’ in English, and the plurality inference would arise by the competition with ‘quién’_{SG} in the relevant contexts. As long as this inference is an implicated presupposition, *quiénes*-interrogatives should be available whenever the singular alternative is not, namely in both *more-than-one* and ignorance scenarios. Hence, the distribution of ‘quiénes’ should be strictly equivalent to the one of all the other plural interrogatives, headed by ‘cuáles’ or ‘qué’ in Spanish or by ‘which’ in English. In (8b), we already showed that *quiénes*-interrogatives are deviant in ignorance scenarios.⁸ This pattern persists across different contexts, contrasting –crucially– with the behaviour of plural *which*-questions in English and with *cuáles*-questions in Spanish:

- (11) *Juan is expecting at least one of his friends to go to the party.*
- a. Juan no sabe todavía quién de sus amigos va a ir a la fiesta.
Juan not know yet who.SG of his friends FUT PREP go to the party.
‘Juan doesn’t know yet who (among his friends) will go to the party.’
 - b. #/?? Juan no sabe todavía quiénes de sus amigos van a ir a la fiesta.
#/?? Juan not know yet who.PL of his friends FUT PREP go to the party.
‘Juan doesn’t know yet who (among his friends) will go to the party.’
 - c. Juan no sabe todavía qué/cuáles amigos van a ir a la fiesta.
Juan not know yet what/which.PL friends FUT PREP go to the party.
‘Juan doesn’t know yet what/which friends will go to the party.’
 - d. Juan doesn’t know yet which friends will go to the party.

Plural interrogatives headed by ‘which’ in English and by ‘cuáles’ or ‘qué’ in Spanish can be felicitously uttered in ignorance contexts, as observed in (11c) and (11d). This indicates that the *plurality inference* for these interrogatives is indeed an implicated presupposition: the relevant questions can be used as soon as the presupposition of their singular counterpart is not satisfied, including in an ignorance scenario. Instead, *quiénes*-interrogatives are still incompatible with this context (e.g., (11b)), suggesting that the plurality inference for these questions is not an implicated presupposition.

A further contrast between *quiénes*-questions, on the one hand, and plural *which* and *cuáles* questions, on the other, is obtained when these interrogatives are embedded under a universal quantificational adverb (e.g. ‘each day of the week’). Consider the scenario described in (12). This ‘mixed’ context ensures that every day there was *at least* one person the professor received. As predicted by different accounts of plurality (see Križ, 2015; Spector, 2007a; Zweig, 2009 for a discussion of this kind of examples), sentences such as (12a) and (12b) can be considered true descriptions of such situation (i.e. the professor received one or more than one student each day of the week’). Instead, *quiénes*-interrogatives are perceived as deviant by Spanish native speakers.

⁸ Arguably, the sluicing example provided in (8b) is also deviant when we replace ‘quiénes’ by ‘which ones’ in English or by ‘cuáles’ in Spanish (e.g. i). However, the singular alternatives in (ii) are even more degraded in both languages.

- a. (i) ? At least one of my clients called yesterday, but I am not sure which ones.
(ii) #At least one of my clients called yesterday, but I am not sure which one.
- b. (i) ?Al menos uno de mis clientes llamó ayer, pero no estoy segura (de) cuáles.
(ii) #Al menos uno de mis clientes llamó ayer, pero no estoy segura (de) cuál.

COMMON GROUND		
<i>Exactly one person VP</i>	<i>At least one person VP</i>	<i>More than one person VP</i>
Which/What person VP? (EN) Cuál/Qué persona VP? (SP)	Which/What people VP? (EN) Cuáles/Qué personas VP? (SP)	
	Who VP? (EN) Quién VP? (SP)	Quiénes VP? (SP)

Table 10.1: Distribution of interrogatives depending on the contexts where they can appear. Shaded cells indicate the alternatives that carry stronger presuppositions.

- (12) *The professor received one student every day and at least one day he received multiple students.*
- The professor knows which students came to his office each day of the week.
 - El profesor sabe qué/cuáles estudiantes vinieron a su oficina cada día de la semana.
The professor knows what/which.PL students came to his office each day of the week.
'The professor knows which students came to his office each day of the week.'
 - ?? El profesor sabe quiénes de sus estudiantes vinieron a su oficina cada día de la semana.
?? The professor knows who.PL of his students came to his office each day of the week.
'The professor knows who (among his students) came to his office each day of the week.'

Note that this contrast cannot be explained by relying on relevance or contextual differences (Križ, 2015): for each of the three sentences in (12) it is equally irrelevant whether one or more than one student was received by the professor.

The examples presented in (8)-(12) (see illustration in Table 10.1) suggest that *quién* and *quiénes* interrogatives cannot be explained with the same tools we used to account for *which* and *cuál* interrogatives. Namely, they cannot be explained in terms of weak account of plurality. Rather, the properties of these questions can be nicely integrated within a *strong plural semantics* (Chierchia, 1998; de Swart & Farkas, 2010), such that the plural 'quiénes' carries a *more-than-one* requirement as part of its literal meaning. In principle, we cannot commit to including such requirement directly in the semantics of the plural morpheme: a weak plural meaning is still required in order to explain the behaviour of the different plural interrogatives in both Spanish and English.

- (13) $\llbracket \text{QUIÉNES}^{[+WH]} \rrbracket^w = \lambda F_{et}. \exists x. x \in \llbracket \text{people} \rrbracket^w \wedge \underline{|x|} > 1 \wedge F(w)(x)$, where $\llbracket \text{people} \rrbracket^w$ is closed under sum.

As we will observe in more detail in Section 10.4, a lexical entry such as (13) ensures that the denotation of *quiénes*-interrogatives includes plural propositions, making them only compatible with *more-than-one* contexts. In this respect, *quiénes*-interrogatives are the mirror image of singular questions headed by 'cuál' and 'which': while the latter presuppose that the complete answer involves an atomic individual, the former presuppose that the complete answer involves a non-atomic individual.

10.3.2 D-linking requirement

We have proposed that *quiénes*-interrogatives impose a *more-than-one* requirement into the common ground. While satisfying this requirement is a necessary condition to license a *quiénes*-interrogative, it is not a sufficient one. Additionally, the plural ‘*quiénes*’ seems to require a *discourse-linked* context, where the domain of individuals over which the quantifier ranges is either contextually salient or has been previously introduced into the discourse (Comorovski, 2013; Pesetsky, 1987).

Consider the context in (14). This scenario ensures that the speaker believes that more than one person is in the house, but differs from previous examples in that the individuals contained in the set of people are not contextually salient or discourse-given. *Quiénes*-interrogatives are not licensed in these cases, and only the alternative with ‘*quién*’ can be uttered (14a). A similar pattern is attested with plural indefinites in (14b) (Alonso-Ovalle & Menéndez-Benito, 2011; Etxeberria & Giannakidou, 2010).⁹ As expected, the contrast is reversed if an antecedent for the *wh*-phrase has been previously introduced in the discourse (14c).¹⁰

- (14) *Mary and John arrive at their apartment, where there is supposed to be no one. They hear two people whispering inside. Mary asks:*
- a. *Quién (#quiénes) está ahí?*
Who.SG (#who.PL) is there?
‘Who is in there?’
 - b. *Algún ladrón (#algunos ladrones) debe haber entrado.*
Some.SG thief (#some.PL thieves) must have entered.
‘Some thief must have broken-in’
 - c. *Dos personas están hablando en el dormitorio, pero no sé quiénes*
Two people are talking in the bedroom, but not know who.PL
(#quién) son.
(#who.SG) are.
‘Two people are talking in the bedroom, but I don’t know who they are.’

Another piece of evidence indicating the need of an additional *d-linking* requirement comes from the possibility of combining the singular ‘*quién*’ with collective and reciprocal predicates in (15a) and (15b). Collective and reciprocal predicates can *only* be predicated of pluralities; therefore, the speaker is *by default* opinionated about the cardinality of the expected answer.

⁹An anonymous reviewer points out that the contrast in (14a) is likely to have a different source from the one in (14b). Indeed, the singular ‘*algún*’ is an ignorance indefinite, whereas the plural ‘*algunos*’ is not, explaining why only the former is acceptable in this type of scenarios. In contrast, ‘*quién*’ does not introduce ignorance—at least not in the same sense.

¹⁰‘*Quién*’ and ‘*quiénes*’ also differ in their ability to appear in interrogatives without existential import. As *who*-interrogatives in English (Krifka, 2011), singular *quién*-interrogatives can appear both in *there*-insertion contexts (e.g. i) and in scenarios where the speaker puts in doubt the existence of propositions in the set of true answers (e.g. rhetorical questions in (ii); Aguero-Bautista, 2001). *Quiénes*-interrogatives are deviant in these contexts.

- (i) *Quién (# quiénes) hay en la fiesta?*
Who.SG (# who.PL) was at the party?
‘Who was there at the party?’
- (ii) *Mary’s parents decided not to let her go out on Friday’s night. Very upset, Mary says to them:*
Quién (# quiénes) se creen que son?
Who.SG (# who.PL) REFL believe that are?
‘Who do you think you are?’

The availability of appearing in *there*-insertion contexts has been traditionally used to place DPs in the (in)definiteness scale (Heim, 1987). In (10), ‘*quién*’ proves to be better than ‘*quiénes*’ in this context, suggesting that the later is a stronger DP than the former. Establishing a connection between *d-linking*, on the one hand, and definiteness and existential import, on the other, might help capturing all these contrasts with an unique account.

Given that *quiénes*-interrogatives have stronger presuppositions than their singular alternative (cf. (13)), questions such as (15a) should be blocked. Nonetheless, these *quién*-interrogatives are actually preferred in *non-d-linked* contexts. A similar pattern arises when quantificational adverbs quantify over the embedded question (Quantificational Variability Effects, Berman, 1991). The sentence in (15c) triggers the inference that multiple people called and for most of them, Mary knows that they called. The embedded interrogative ‘quién llamó’ (‘who_{SG} called’) needs to have a plural maximal answer in order to allow such reading.

- (15) a. Quién se juntó ayer a la noche?
Who.SG REFL gathered yesterday at the night?
‘Who gathered last night?’
- b. Quién se conoce entre sí en la fiesta?
Who.SG REFL know between them at the party?
‘Who knows each other at the party?’
- c. María sabe en su mayoría quién llamó.
María knows in its majority who.SG called
‘Mary mostly knows who called’

The examples above suggest that the plural ‘quiénes’ can only range over a discourse or contextually salient set: ‘quiénes’ is always understood as ‘quiénes out of the salient set of people’.¹¹ This property, described as *d-linking*, has been used to account for the syntactic behaviour of interrogative sentences (e.g. *which*-interrogatives in English). However, it has not been formally defined in the semantics literature, being alternatively assimilated to specificity or definiteness. For the purposes of this paper, I will define *d-linking* as follows:

- (16) (i) An interrogative phrase is *d-linked* in a context *c* if its restrictor *P* is understood to denote a contextually salient set of individuals that are *P* and that can be referred to by a pronoun or by the definite description ‘The *Ps*’. In such a case, the interrogative phrase can be paraphrased by a partitive *wh*-phrase of the form ‘Which of the *Ps*’.¹²
- (ii) An interrogative phrase has a *d-linking requirement* if it cannot be felicitously used in a context *c* unless it is *d-linked* in *c*.

In scenarios such as (14), neither the pronoun ‘they’ nor the definite description ‘the thieves’ are licensed, indicating that the context does not provide a “contextually salient” set of thieves in the sense of (16)i. In other words, although there might be a sense in which *there is* a salient set of two people whispering inside the apartment (as suggested by an anonymous reviewer), the interrogative cannot be *d-linked* in this context.

- (17) (cf. context in (14))
- a. #Ellos deben ser ladrones.
‘They must be thieves’
- b. #Los ladrones nos están robando.
‘The thieves must have broken in’
- c. #Cuáles de los ladrones entraron?

¹¹One might be tempted to treat the interrogative ‘quiénes’ as a definite description (e.g. *the*(λx *people*(*x*))), in the same way it has been proposed for *which*-phrases in English (Novel & Romero, 2010; Rullmann & Beck, 1998). The *d-linking* requirement would then be captured by the standard familiarity constraint for definite phrases (i.e. the reference of the definite has to be “familiar” to the audience, Heim, 1983). We will leave this possibility open for future research.

¹²The notion of “contextually salient individual” used here does not in fact cover cases where, informally speaking, a *discourse-referent* (to use DRT terminology, Heim, 1983; Kamp, Van Genabith, & Reyle, 2011) is introduced by indefinites in preceding discourse, allowing for subsequent pronouns to pick up that referent. Whereas contexts that license pronouns also license ‘quiénes’, a full formalization of the notion of “contextually salient individual” would need to rely on a dynamic semantics framework. That, however, goes beyond the scope of this paper.

‘Which of the thieves have broken in?’

Some interrogative elements have a *d-linking requirement*: they can only be uttered in a context where they are d-linked. Others, instead, are underspecified for d-linking, and they can appear in any context. *Wh*-words such as ‘which’ and ‘quiénes’ seem to belong to the first group, whereas English ‘who’ and Spanish ‘quién’ belong to the second. Interestingly, some of these interrogative elements can optionally take as restrictors overt partitives that contain pronouns or definite descriptions. Whenever they do, we can assume that the utterance context provides a contextual reference for the overt partitive, and therefore the context is d-linked for the restrictor. This is the case for ‘quién’ and ‘quiénes’: they can both take overt partitives (e.g. (18)).

- (18) a. Quién de tus amigos llamó ayer?
 who.SG of your friends called yesterday?
 ‘Which one of your friends called yesterday?’
 b. Quiénes de tus amigos llamaron ayer?
 who.PL of your friends called yesterday?
 ‘Which ones of your friends called yesterday?’

We can thus predict that whenever the *wh*-element takes a partitive and the cardinality requirement is met (i.e. *more-than-one* contexts), only the alternative with ‘quiénes’ will be licensed.¹³ This is indeed the case:

- (19) a. * Quién de tus amigos se juntó ayer?
 * who.SG of your friends REFL gathered yesterday?
 ‘Which one of your friends gathered yesterday?’
 b. Quiénes de tus amigos se juntaron ayer?
 who.PL of your friends REFL gathered yesterday?
 ‘Which ones of your friends gathered yesterday?’
 c. * María sabe en su mayoría quién de sus amigos llamó.
 * María knows in its majority who.SG of her friends called.
 ‘Mary mostly knows which one of her friends called.’
 d. María sabe en su mayoría quiénes de sus amigos llamaron.
 María knows in its majority who.PL of her friends called.
 ‘Mary mostly knows which ones of her friends called.’

Before concluding this section, let me briefly note that the contrasts between ‘quién’ and ‘quiénes’ discussed in Section 10.3.1 are expected to become sharper whenever it is guaranteed that the interrogative is d-linked in the context; namely, whenever the *wh*-word takes an overt partitive.

10.4 Account

The use conditions of *quién* and *quiénes*-interrogatives can be summarized as follows (see also Table 10.2 for a complete paradigm):

¹³For inherently d-linked words, having an overt partitive does not make any difference. For example, ‘which’ is one of the few *wh*-words in English that can take an overt partitive as a complement. However, since it is also inherently d-linked, both *which*-interrogatives below can only be felicitously uttered if the context provides a salient set of books, from where the speaker is asking for a choice.

- (i) Which book did you read?
 (ii) Which of the books did you read?

- (20) A sentence of the form *Quiénes VP?* can be felicitously uttered in *c* iff:
- a. **cardinality requirement**
the question presupposes its complete answer involves a non-atomic individual: $c \cap \llbracket \text{more than one person } VP \rrbracket$
 - b. **d-linking requirement**
the question presupposes its complete answer will be drawn from a ‘contextually salient’ (or previously introduced) set of individuals.

Otherwise, the alternative *Quién VP?* should be used.

The plural ‘*quiénes*’ is constrained by the conjunction of (a) and (b): As long as one of the two requirement is not fulfilled in the common ground, the alternative with ‘*quién*’ will be the only one available option. *Quién*-interrogatives are therefore underspecified, and their use conditions are disjunctive: they can be used either when the plurality requirement is not met or when the context is not d-linked.

	COMMON GROUND		
	<i>Exactly one person VP</i>	<i>At least one person VP</i>	<i>More than one person VP</i>
Interrogative is <i>d-linked</i> in the context	Which person <i>VP?</i> (EN) Cuál/Qué persona <i>VP?</i> (SP)	Which people <i>VP?</i> (EN) Cuáles/Qué personas <i>VP?</i> (SP)	
	Quién <i>VP?</i> (SP)		Quiénes <i>VP?</i> (SP)
Interrogative is not <i>d-linked</i> in the context	What person <i>VP?</i> (EN) Qué persona <i>VP?</i> (SP)	What people <i>VP?</i> (EN) Qué personas <i>VP?</i> (SP)	
	Who <i>VP?</i> (EN) Quién <i>VP?</i> (SP)		

Table 10.2: Final Distribution of interrogatives depending on the contexts where they can appear.

Given that the two requirements in (20) are only imposed on ‘*quiénes*’, it is natural to infer that they are encoded in the semantics of plural morphology. However, in light of the broader set of data (e.g. *which* and *cuál*-interrogatives), this assumption appears faulty. While the exact contribution of plural marking needs to be further explored, in what follows we will assume that the plurality and d-linking requirements emerge specifically for ‘*quiénes*’, as part of its lexical meaning.

Let us now inspect our account by working on the examples in (4). Since providing a formal definition of d-linking exceeds the scope of this paper, the requirement in (20b) will be expressed by simply assuming that ‘*quiénes*’ includes a pronominal element, represented by an index whose reference is determined by the assignment function and serves as the restrictor of the interrogative quantifier. Following Heim and Kratzer’s (1998) account of referring pronouns, an appropriated context for the use of ‘*quiénes_i*’ is one that provides a specific assignment function for the index *i* (i.e. *appropriated condition*). Moreover, ‘*quiénes_i*’ triggers the presupposition that the value assigned to *i* by the assignment function is contained in the set of people, which contains atoms as well as pluralities. This presupposition is cashed out as a definiteness condition in (21b), given Heim and Kratzer’s notational conventions. The existential quantifier will range over the referents introduced by this anaphor; hence, ‘*quiénes_i*’ can be thought as equivalent to ‘which of them’, modulo the plurality requirement. For the sake of the example, we will assume the d-linking requirement is met in the utterance context (i.e. the interrogative phrase is d-linked in the context). That is to say, the utterance context provides a ‘contextually salient set of individuals’, which in our example are Mary, John and Bill.

Adopting the lexical entries in (21), the interrogatives in (4a) and (4b) will have the denota-

tions in (22a) and (22b). Note that the LFs in (22) are derived by means of Heim and von Stechow's (2016) take on Karttunen (1977) semantics, as we have done before.

- (21) a. $\llbracket \text{QUIÉN}^{[+WH]} \rrbracket^w = \lambda F_{et} \exists x. x \in \llbracket \text{people} \rrbracket^w \wedge F(x)(w)$
 b. $\llbracket \text{QUIÉNES}_i^{[+wh]} \rrbracket^{w,g} = \lambda F_{et} : g(i) \in \llbracket \text{people} \rrbracket^w . \exists x. x \preceq g(i) \wedge |x| > 1 \wedge F(w)(x)$
- (22) $\llbracket \text{people} \rrbracket = \{m, j, b, m \oplus j, m \oplus b, j \oplus b, m \oplus j \oplus b\}$
- a. $\llbracket (4a) \rrbracket^w = \lambda p. \exists x. x \in \llbracket \text{people} \rrbracket^w \wedge p = \lambda w'. x \text{ called in } w' = \left\{ \begin{array}{l} \lambda w'. m \text{ called in } w', \\ \lambda w'. b \text{ called in } w', \\ \lambda w'. j \text{ called in } w', \\ \lambda w'. j \oplus m \text{ called in } w', \\ \lambda w'. j \oplus b \text{ called in } w', \\ \lambda w'. b \oplus m \text{ called in } w', \\ \lambda w'. j \oplus m \oplus b \text{ called in } w' \end{array} \right\}$
- b. $\llbracket (4b) \rrbracket^{w,g} = \lambda p : g(i) \in \llbracket \text{people} \rrbracket^w . \exists x. x \preceq g(i) \wedge |x| > 1 \wedge p = \lambda w'. x \text{ called in } w' = \left\{ \begin{array}{l} \lambda w'. j \oplus m \text{ called in } w', \\ \lambda w'. j \oplus b \text{ called in } w', \\ \lambda w'. b \oplus m \text{ called in } w', \\ \lambda w'. j \oplus m \oplus b \text{ called in } w' \end{array} \right\}$

After applying the ANS operator to (22a) and (22b), the two corresponding LFs have different presuppositional strength: the set of worlds where (22b) is defined is included in set the worlds (22a) is defined. In every context where it is common ground that exactly one person went to the party, (22b) will yield to a presupposition failure, leading to the attested oddness (e.g. (8a)). Conversely, if it is common ground that more than one person (out of a salient set of people) went to the party, the two interrogatives will be contextually equivalent.¹⁴ A principle such as *Maximize Presupposition!* will then select (22b), since this is the LF carrying a stronger presupposition.

When a *quién*-interrogative is uttered in a context where the d-linking requirement is met, one should infer that it's not presupposed that the question has a plural answer, or the alternative with 'quiénes' should have been used. Questions such as (4a) therefore trigger an ignorance inference: according to what is common knowledge, it is possible that one or more than one person has called. This explains why *quién*-interrogatives are compatible with *ignorance* scenarios. However, if the speaker can be considered to be reliable and knowledgeable, following the logic of the epistemic step for implicated presuppositions or anti-presuppositions, an 'exactly-one' inference (cf. *strengthened* implicated presupposition) will be derived. One would then be in a position to infer that, according to what is common ground, exactly one person has called.¹⁵

Finally, whenever the d-linking requirement is not fulfilled in the common ground (i.e. there is no reference for the anaphora in the context), *quiénes*-interrogatives are predicted to yield to a presupposition failure. The denotation of the question will be empty, and the existential presupposition of the ANS operator will not be satisfied in the common ground. *Quién*-interrogatives will then be the only available alternative. Note that the two interrogatives can only be contextually equivalent (7a) if and only if the context both allows the interrogative to be d-linked and involves more than one person who VP. Since in contexts where the *d-linking* requirement is not met *Maximize Presupposition!* can never be applied, in these contexts *quién*-questions will not trigger any inference regarding the cardinality of the expected answer (cf. no

¹⁴Note that these two interrogatives are only contextually equivalent if we compare LFs containing the ANS operator: applying the answerhood operator to these the two interrogatives will return the same maximal true answer for all worlds compatible with the context.

¹⁵At this point, it's important to note a difference of strength between implicated presuppositions associated to plural *which*-interrogatives (cf. Section 2) and to *quién*-interrogatives. As observed, in most contexts, *which*-interrogatives (e.g. (1b)) trigger a quite *strong* plurality inference (i.e. it is common ground that more than one client called). Instead, singular *quién*-questions tend to convey complete ignorance, compatible with a standard implicated presupposition (i.e. it is possible, according to common knowledge, that only one person have had called).

implicated presupposition). This explains the distribution attested in examples such as (14) and (15).

10.5 Beyond interrogatives: Conclusions and issues for further research

The account provided in this article contributes to the current research on number semantics by bringing interrogative quantifiers into the picture. We have addressed the contrast between *quién* and *quiénes*-interrogatives in Spanish, claiming that plural morphology has a *strong* semantic import into the quantifier meaning, whereas singular marking is semantically vacuous (i.e. *strong plural/weak singular* semantics). We have then argued that the distribution of the two interrogatives arises from the existence of two requirements (cardinality and d-linking) which seem to be linked to plural morphology.

As a result, our analysis shows that current approaches to number marking need to be made more sophisticated to account for both cross-linguistic and within-language variation. Although the behavior of ‘quién’ and ‘quiénes’ is well explained by a *strong plural/weak singular* semantics, the meaning of other indefinites in both Spanish and English is still best accounted for in terms of a weak semantics for plural morphology. As observed, singular *which* and *cuál* questions do trigger a uniqueness presupposition (e.g. (1a)/(2a)), whereas plural *which* and *cuáles* interrogatives are fully compatible with ignorance contexts, where it’s not common knowledge that exactly one individual will be named in the complete answer (e.g. (11)). Accounts relying on a unified semantics for singular and plural morphemes (cf. Sauerland, 2003; Sauerland et al., 2005) are therefore shown to be incorrect.

Rather, one might suggest that number morphology has different semantic import depending on whether it applies to single-word quantifiers (i.e. often denominated indefinite pronouns or generalized quantifiers) or to determiners. Specifically, traditional approaches could account for number marking applied to determiners, whereas single-word quantifiers would follow a weak singular/strong plural pattern. An analysis along these lines would explain the difference between ‘quién’/‘quiénes’ and ‘cuál’/‘qué’-phrases in Spanish (and in English). Indeed, a contrast between these two different kinds of quantifiers is also attested in the declarative domain, where some singular single-word quantifiers seem to have a weak semantics, making them compatible with a collective predicate (p.c. Benjamin Spector). An illustration of such behaviour is shown in (23) for different quantifiers in English. The same pattern is attested in Spanish.

- (23) a. Everyone gathered
b. *Every student gathered
- (24) a. Nobody gathered
b. *No student gathered

Note that, if the ideas sketched above are on track, every single-word quantifier that can receive both plural and singular marking is predicted to have the same distribution as ‘quién’ and ‘quiénes’ in Spanish.

There are, however, some counterexamples to this generalisation. A first notable exception is the distribution of English ‘someone’. Like the singular determiner ‘some’, ‘someone’ is unable to appear as the subject of a collective predicate (e.g. **Someone gathered*), suggesting that it is not a weak singular despite being a single-word quantifier. Additionally, some plural determiners might also impose *strong* plural requirements, even independently of having a strong singular as their alternative. This seems to be the case for the existential determiner ‘unos’ (*ones*) in Spanish, which directly competes with the singular indefinite ‘un’ (*a*) (see Alonso-Ovalle & Menéndez-Benito, 2011; Martí, 2008 for the relevant examples).

All these ideas are tentative, but they serve to illustrate the complexity of the problem, reinforcing the need for a “mixed” approach to number semantics in order to account for the complete range of data. Further research will hopefully help to clarify this issue.

Let us finally point out that, besides discussing the cardinality requirements associated to number morphology, our account of *quién* and *quiénes*-interrogatives also establishes an unexpected link between plural marking and d-linking. Although an homologous connection has been attested for plural epistemic indefinites in Spanish (Alonso-Ovalle & Menéndez-Benito, 2011), as illustrated in (14), the details of such relation remain unexplained in our account. We remain agnostic as to whether d-linking is a requirement directly triggered by plural morphology when it applies to certain quantifiers.

References

- Aguero-Bautista, C. (2001). *Cyclicity and the scope of wh-phrases* (Unpublished doctoral dissertation).
- Alonso-Ovalle, L., & Menéndez-Benito, P. (2011). Domain restrictions, modal implicatures and plurality: Spanish algunos. *Journal of Semantics*, 28(2), 211–240. doi: 10.1093/jos/ffq016
- Anderson, C. (2004). *The structure and real-time comprehension of quantifier scope ambiguity* (Unpublished doctoral dissertation). Northwestern University Evanston, IL.
- 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(4), 390–412.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, 68(3), 255–278.
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using eigen and s4 [Computer software manual]. Retrieved from <http://CRAN.R-project.org/package=lme4> (R package version 1.1-7)
- Beghelli, F., & Stowell, T. (1997). Distributivity and negation: The syntax of each and every. In *Ways of scope taking* (pp. 71–107). Springer.
- Berman, S. R. (1991). On the semantics and logical form of wh-clauses.
- Bock, K. (1986). Syntactic persistence in language production. *Cognitive Psychology*, 18(3), 355–387.
- Bock, K., & Loebell, H. (1990). Framing sentences. *Cognition*, 35(1), 1–39.
- Bott, L., & Chemla, E. (2016). Shared and distinct mechanisms in deriving linguistic enrichment. *Journal of Memory and Language*, 91, 117–140.
- Bott, O., & Radó, J. (2009). How to provide exactly one interpretation for every sentence, or what eye movements reveal about quantifier scope. In *The fruits of empirical linguistics* (Vol. 1, pp. 1–24).
- Branigan, H., & Messenger, K. (2016). Consistent and cumulative effects of syntactic experience in children’s sentence production: Evidence for error-based implicit learning. *Cognition*, 157, 250–256.
- Branigan, H., & Pickering, M. (2017). Structural priming and the representation of language. *Behavioral and Brain Sciences*, 40.
- Brasoveanu, A., & Dotlačil, J. (2015). Processing pluralities: Syntax and the lexicon. *CUNY 28*. Retrieved from https://people.ucsc.edu/~abrsvn/CUNY_poster_2015.pdf
- Brisson, C. M. (1998). *Distributivity, maximality, and floating quantifiers* (Unpublished doctoral dissertation). Citeseer.
- Brooks, P. J., & Braine, M. D. (1996). What do children know about the universal quantifiers all and each? *Cognition*, 60(3), 235–268.
- Champollion, L. (2010). Parts of a Whole: Distributivity as Bridge between Aspect and Measurement. , 1–10.
- Champollion, L. (2014). Covert distributivity in algebraic event semantics. *lingbuzz/002097*.
- Champollion, L. (to appear). Distributivity, collectivity, and cumulativity. In L. Matthewson,

- C. Meier, H. Rullmann, & E. Zimmermann (Eds.), *Wiley's companion to semantics*. Wiley.
- Champollion, L., & Krifka, M. (2014). to appear. mereology. *Cambridge handbook of semantics*. Cambridge University Press. <http://ling.auf.net/lingbuzz/002099>.
- Chang, F., Dell, G. S., & Bock, K. (2006). Becoming syntactic. *Psychological Review*, 113(2), 234.
- Chemla, E. (2008). An Epistemic Step for Anti-Presupposition. *Journal of Semantics*, 25(2), 141–173. Retrieved from <http://www.emmanuel.chemla.free.fr/Material/AntiPresuppositionAndEpistemicStep.pdf> doi: 10.1093/jos/ffm017
- Chemla, E., & Bott, L. (2014). Processing inferences at the semantics/pragmatics frontier: Disjunctions and free choice. *Cognition*, 130(3), 380–396.
- Chemla, E., & Bott, L. (2015). Using structural priming to study scopal representations and operations. *Linguistic Inquiry*.
- Chierchia, G. (1993). Questions with quantifiers. *Natural language semantics*, 1(2), 181–234.
- Chierchia, G. (1998). Plurality of mass nouns and the notion of 'semantic parameter'.
- Chierchia, G., Fox, D., & Spector, B. (2012). The grammatical view of scalar implicatures and the relationship between syntax and semantics. *Semantics: An international handbook of natural language meaning*, 1.
- Chomsky, N. (1976). *Conditions on rules of grammar*.
- Chomsky, N., Belletti, A., & Rizzi, L. (2002). An interview on minimalism. *N. Chomsky, On Nature and Language*, 92–161.
- Clark, D. G., & Kar, J. (2011). Bias of quantifier scope interpretation is attenuated in normal aging and semantic dementia. *Journal of Neurolinguistics*, 24(4), 401–419.
- Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive psychology*, 3(3), 472–517.
- Comorovski, I. (2013). *Interrogative phrases and the syntax-semantics interface* (Vol. 59). Springer Science & Business Media.
- Cooper, R. (2013). *Quantification and syntactic theory* (Vol. 21). Springer Science & Business Media.
- Cranford, E. A., & Moss, J. (2017). Mouse-tracking evidence for parallel anticipatory option evaluation. *Cognitive processing*, 1–24.
- Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford University Press on Demand.
- Dale, R., & Duran, N. D. (2011). The Cognitive Dynamics of Negated Sentence Verification. *Cognitive Science*, 35, 983–996. doi: 10.1111/j.1551-6709.2010.01164.x
- Dayal, V. (1996). *Locality in wh quantification: questions and relative clauses in hindi*. Kluwer Academic Publishers.
- Degen, J. (2013). *Alternatives in pragmatic reasoning* (Unpublished doctoral dissertation). University of Rochester.
- de Swart, H., & Farkas, D. (2010). The semantics and pragmatics of plurals. *Semantics and pragmatics*, 3, 6–1.
- Donders, F. C. (1969). On the speed of mental processes. *Acta psychologica*, 30, 412–431.
- Dotlacil, J. (2010). *Anaphora and distributivity: A study of same, different, reciprocals and others*. Netherlands Graduate School of Linguistics.
- Dowty, D., et al. (1987). Collective predicates, distributive predicates and all. In *Proceedings of the 3rd escol* (pp. 97–115).
- Etxeberria, U., & Giannakidou, A. (2010). Contextual domain restriction and the definite determiner. *Context-dependence, perspective and relativity*, 6.
- Farmer, T. a., Cargill, S. a., Hindy, N. C., Dale, R., & Spivey, M. J. (2007). Tracking the continuity of language comprehension: computer mouse trajectories suggest parallel syntactic processing. *Cognitive science*, 31, 889–909. doi: 10.1080/03640210701530797
- Feiman, R., & Snedeker, J. (2016). The logic in language: How all quantifiers are alike, but each quantifier is different. *Cognitive Psychology*, 87, 29–52.
- Ferreira, V. S., & Bock, K. (2006). The functions of structural priming. *Language and cognitive*

- processes*, 21(7-8), 1011–1029.
- Filcik, R., Paterson, K. B., & Liversedge, S. P. (2004). Processing doubly quantified sentences: Evidence from eye movements. *Psychonomic Bulletin & Review*, 11(5), 953–959.
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS one*, 8(10).
- Fodor, J. D. (1982). The mental representation of quantifiers. In *Processes, beliefs, and questions* (pp. 129–164). Springer.
- Fodor, J. D., & Sag, I. A. (1982). Referential and quantificational indefinites. *Linguistics and philosophy*, 5(3), 355–398.
- Fox, D. (2000). *Economy and semantic interpretation* (Vol. 35). MIT press.
- Fox, D. (2003). On logical form. *Minimalist syntax*, 82–123.
- Frazier, L. (1999). *On sentence interpretation* (Vol. 22). Springer Science & Business Media.
- Frazier, L., Pacht, J. M., & Rayner, K. (1999). Taking on semantic commitments, ii: collective versus distributive readings. *Cognition*, 70(1), 87–104.
- Freeman, J. B., & Ambady, N. (2010). MouseTracker: software for studying real-time mental processing using a computer mouse-tracking method. *Behavior research methods*, 42(1), 226–241. doi: 10.3758/BRM.42.1.226
- Freeman, J. B., Dale, R., & Farmer, T. A. (2011). Hand in motion reveals mind in motion. *Frontiers in Psychology*, 2(APR), 1–6. doi: 10.3389/fpsyg.2011.00059
- Freeman, J. B., & Johnson, K. L. (2016). More than meets the eye: split-second social perception. *Trends in cognitive sciences*, 20(5), 362–374.
- Frege, G. (1891). Function and concept. translated in m. black and p. geach. *Translations of Philosophical Writings of Gottlob Frege*, 56–78.
- Gil, D. (1982). Quantifier scope, linguistic variation, and natural language semantics. *Linguistics and Philosophy*, 5(4), 421–472.
- Gillen, K. (1991). *The comprehension of doubly quantified sentences*. (Unpublished doctoral dissertation). Durham University.
- Gillon, B. S. (2004). Ambiguity, indeterminacy, deixis and vagueness: evidence and theory.
- Glass, L. (n.d.). *An empirical investigation of distributivity, lexical semantics, & world knowledge*. (Presented at Linguistic Evidence 2018)
- Glass, L. (2018). Deriving the distributivity potential of adjectives via measurement theory. *Proceedings of the Linguistic Society of America*, 3(1), 49–1-14.
- Hagstrom, P. (2003). What questions mean. *Glott International*, 7(7/8), 188–201.
- Hamblin, C. L. (1973). Questions in montague english. *Foundations of language*, 10(1), 41–53.
- Hartsuiker, R. J., & Kolk, H. H. (1998). Syntactic facilitation in agrammatic sentence production. *Brain and Language*, 62(2), 221–254.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference and prediction* (Second Edition ed.). New York: Springer.
- Helman, E., Stoller, R. M., & Freeman, J. B. (2014). Advanced mouse-tracking analytic techniques for enhancing psychological science. *Group Processes & Intergroup Relations*, 1–18. doi: 10.1177/1368430214538325
- Heim, I. (1983). File change semantics and the familiarity theory of definiteness.
- Heim, I. (1987). Where does the definiteness restriction apply? evidence from the definiteness of variables. *The representation of (in) definiteness*, 14.
- Heim, I. (1991). Articles and definiteness. *Semantics. An international handbook of contemporary research*. Berlin: De Gruyter.
- Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar* blackwell oxford google scholar.
- Heim, I., & von Stechow, K. (2016). *Lecture notes in advanced semantics*. (MIT)
- Hornstein, N. (1984). *Logic as grammar*. MIT Press.
- Huang, Y. T., Spelke, E., & Snedeker, J. (2013). What exactly do numbers mean? *Language*

- Learning and Development*, 9(2), 105–129.
- Ioup, G. (1995). Some universals for quantifier scope. In J. Kimball (Ed.), *Syntax and semantics 4* (pp. 37–58). New York: Academic Press.
- Jackendoff, R. S. (1972). Semantic interpretation in generative grammar.
- Jaeger, T. F. (2008). Categorical data analysis: Away from anovas (transformation or not) and towards logit mixed models. *Journal of memory and language*, 59(4), 434–446.
- Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology*, 61(1), 23–62.
- Jaeger, T. F., & Snider, N. E. (2013). Alignment as a consequence of expectation adaptation: Syntactic priming is affected by the prime's prediction error given both prior and recent experience. *Cognition*, 127(1), 57–83.
- Johnson-Laird, P. (1969). On understanding logically complex sentences. *The Quarterly Journal of Experimental Psychology*, 21(1), 1–13.
- Kamp, H., Van Genabith, J., & Reyle, U. (2011). Discourse representation theory. In *Handbook of philosophical logic* (pp. 125–394). Springer.
- Karttunen, L. (1977). Syntax and semantics of questions. *Linguistics and philosophy*, 1(1), 3–44.
- Kaup, B., Kelter, S., & Habel, C. (2002). Representing referents of plural expressions and resolving plural anaphors. *Language and Cognitive Processes*, 17(4), 405–450.
- Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Lüdtke, J. (2007). Experiential simulations of negated text information. *The Quarterly Journal of Experimental Psychology*, 60(7), 976–990.
- Kennedy, C. (1999). *Projecting the adjective: The syntax and semantics of gradability and comparison*. Routledge.
- Kratzer, A. (2007). On the Plurality of Verbs. *Event Structures in Linguistic Form and Interpretation* (June 2005), 269–300.
- Krifka, M. (1992). Thematic relations as links between nominal reference and temporal constitution. In I. Sag & A. Szabolsci (Eds.), *Lexical matters* (pp. 29–53). Stanford University: CSLI.
- Krifka, M. (2011). Questions. *Semantics: An international handbook of Natural Language Meaning*, 2, 1742–1785.
- Križ, M. (2015). *Aspects of homogeneity in the semantics of natural language* (Unpublished doctoral dissertation). University of Vienna.
- Kuperberg, G. R., & Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Language, Cognition and Neuroscience*, 31(1), 32–59.
- Kurtzman, H. S., & MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition*, 48(3), 243–279.
- Landman, F. (1996). Plurality. In S. Lappin (Ed.), *The handbook of contemporary semantic theory* (pp. 425–458). Blackwell.
- Landman, F. (2000). *Events and plurality*. Dordrecht: Kluwer.
- Lasnik, P. (1989). On the readings of plural noun phrases. *Linguistic inquiry*, 20(1), 130–134.
- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3), 1126–1177.
- Lidz, J., & Musolino, J. (2002). Children's command of quantification. *Cognition*, 84(2), 113–154.
- Link, G. (1983). The logical analysis of plurals and mass terms: A lattice-theoretical approach. In R. Bäuerle, C. Schwarze, & A. von Stechow (Eds.), *Meaning, use, and interpretation of language* (pp. 303–323). Berlin, Germany: de Gruyter.
- Link, G. (1987). Generalized quantifiers and plurals. In P. Gärdenfors (Ed.), *Generalized quantifiers* (Vol. 31, p. 151–180). Springer Netherlands. Retrieved from http://dx.doi.org/10.1007/978-94-009-3381-1_6 doi: 10.1007/978-94-009-3381-1_6
- Lüdtke, J., Friedrich, C. K., De Filippis, M., & Kaup, B. (2008). Event-related potential correlates of negation in a sentence–picture verification paradigm. *Journal of Cognitive Neuroscience*, 20(8), 1355–1370.

- Maldonado, M., Chemla, E., & Spector, B. (2017). Priming plural ambiguities. *Journal of Memory and Language*, 95, 89 - 101. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0749596X17300098> doi: <http://dx.doi.org/10.1016/j.jml.2017.02.002>
- Maldonado, M., Spector, B., & Chemla, E. (2017). Priming methods in semantics and pragmatics. *Behavioral and Brain Sciences*, 40.
- Martí, L. (2008). The semantics of plural indefinite noun phrases in Spanish and Portuguese. *Natural Language Semantics*, 16(1), 1–37. doi: 10.1007/s11050-007-9023-x
- Marty, P., Chemla, E., & Spector, B. (2013). Interpreting numerals and scalar items under memory load. *Lingua*, 133(June 2011), 152–163. doi: 10.1016/j.lingua.2013.03.006
- Marty, P., Chemla, E., & Spector, B. (2014). Phantom readings: the case of modified numerals. *Language, Cognition and Neuroscience*, 30(4), 462–477.
- May, R. (1977). The grammar of quantification. *Unpublished doctoral dissertation, MIT*.
- May, R. (1985). *Logical form: Its structure and derivation* (Vol. 12). MIT press.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90(2), 227.
- Micham, D. L., Catlin, J., VanDerveer, N. J., & Loveland, K. A. (1980). Lexical and structural cues to quantifier scope relations. *Journal of Psycholinguistic Research*, 9(4), 367–377.
- Milsark, G. L. (1974). *Existential sentences in english* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.
- Montague, R. (1970). English as a formal language.
- Montague, R. (1972). Pragmatics and intensional logic. In *Semantics of natural language* (pp. 142–168). Springer.
- Musolino, J. (1999). Universal grammar and the acquisition of semantic knowledge: An experimental investigation into the acquisition of quantifier-negation interaction in english.
- Musolino, J. (2009). The logical syntax of number words: Theory, acquisition and processing. *Cognition*, 111(1), 24–45.
- Myslin, M., & Levy, R. (2016). Comprehension priming as rational expectation for repetition: Evidence from syntactic processing. *Cognition*, 147, 29–56.
- Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle: An event-related potential study on the pragmatics of negation. *Psychological Science*, 19(12), 1213–1218.
- Novel, M., & Romero, M. (2010). Movement, variables and hamblin semantics..
- Paterson, K. B., Filik, R., & Liversedge, S. P. (2008). Competition during the processing of quantifier scope ambiguities: Evidence from eye movements during reading. *Quarterly Journal of Experimental Psychology*, 61(3), 459–473.
- Patson, N. D. (2014). The processing of plural expressions. *Language and Linguistics Compass*, 8(8), 319–329.
- Patson, N. D., George, G., & Warren, T. (2014). The conceptual representation of number. *The Quarterly Journal of Experimental Psychology*, 67(7), 1349–1365.
- Patson, N. D., & Warren, T. (2010). Evidence for distributivity effects in comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(3), 782.
- Percus, O. (2006). Antipresuppositions. *Theoretical and empirical studies of reference and anaphora: Toward the establishment of generative grammar as an empirical science*, 52, 73.
- Pesetsky, D. (1987). Wh-in-situ: Movement and unselective binding. *The representation of (in) definiteness*, 98, 98–129.
- Piantadosi, S. T., Tily, H., & Gibson, E. (2012). The communicative function of ambiguity in language. *Cognition*, 122(3), 280–291.
- Pickering, M. J., & Ferreira, V. S. (2008). Structural priming: a critical review. *Psychological Bulletin*, 134(3), 427.
- Pylkkänen, L., & McElree, B. (2006). The syntax-semantics interface: On-line composition of

- sentence meaning. In *Handbook of psycholinguistics (second edition)* (pp. 539–579). Elsevier.
- Quine, W. V. O., Churchland, P. S., & Føllesdal, D. (1960). *Word and object*. MIT press.
- R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Raffray, C. N., & Pickering, M. J. (2010). How do people construct logical form during language comprehension? *Psychological Science*, *21*, 1090–1097. doi: 10.1177/0956797610375446
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922.
- Rees, A., & Bott, L. (2018). The role of alternative salience in the derivation of scalar implicatures. *Cognition*, *176*, 1–14.
- Reinhart, T. (1997). Quantifier scope: How labor is divided between qr and choice functions. *Linguistics and philosophy*, *20*(4), 335–397.
- Rett, J. (2008). *Degree modification in natural language*. Rutgers The State University of New Jersey-New Brunswick.
- Roberts, C. (1987). Plural anaphors in distributive contexts. In *Proceedings of the 6th west coast conference on formal linguistics*. stanford: Csl.
- Rullmann, H., & Beck, S. (1998). Presupposition projection and the interpretation of ‘which’-questions. In *Semantics and linguistic theory* (Vol. 8, pp. 215–232).
- Sauerland, U. (2003). A new semantics for number. In *Semantics and linguistic theory* (Vol. 13, pp. 258–275).
- Sauerland, U. (2008). Implicated presuppositions. *The discourse potential of underspecified structures*, 581–600.
- Sauerland, U., Anderssen, J., & Yatsushiro, K. (2005). The plural is semantically unmarked. *Linguistic evidence*(4), 409–430. Retrieved from <http://books.google.com/books?hl=en&lr=&id=tKbrYoZ7zCEC&oi=fnd&pg=PA413&dq=The+Plural+Is+Semantically+Unmarked&ots=cqdl1a5Ww2H&sig=hLOYOMgYnJ94gjj71eM33YvvZsA>
- Sauerland, U., Tamura, A., Koizumi, M., & Tomlinson, J. M. (2015). Tracking down disjunction. In *Jsaí international symposium on artificial intelligence* (pp. 109–121).
- Scha, R. (1981). Distributive, collective and cumulative quantification. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language* (Vol. 2, p. 483–512). Amsterdam: Mathematisch Centrum.
- Scha, R. (1984). Distributive, collective and cumulative quantification. *Truth, Interpretation, and Information*, *GRASS*, *2*, 131–158.
- Scheepers, C. (2003). Syntactic priming of relative clause attachments: Persistence of structural configuration in sentence production. *Cognition*, *89*(3), 179–205.
- Schlotterbeck, F., & Bott, O. (2013). Easy solutions for a hard problem? the computational complexity of reciprocals with quantificational antecedents. *Journal of Logic, Language and Information*, *22*(4), 363–390.
- Schwarz, F. (2013). Maximality and definite plurals – experimental evidence. In E. Chemla, V. Homer, & G. Winterstein (Eds.), *Proceedings of Sinn und Bedeutung 17* (pp. 509–526).
- Schwarzschild, R. (1996). *Pluralities* (Vol. 61). Springer Science & Business Media.
- Schwarzschild, R. (2011). Stubborn distributivity, multiparticipant nouns and the count/mass distinction. In *Proceedings of nels* (Vol. 39, pp. 661–678).
- Scontras, G., & Goodman, N. D. (2017). Resolving uncertainty in plural predication. *Cognition*, *168*, 294–311.
- Sennet, A. (2016). Ambiguity. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* (Spring 2016 ed.). Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2016/entries/ambiguity/>.
- Song, J.-H., & Nakayama, K. (2006). Role of focal attention on latencies and trajectories of visually guided manual pointing. *Journal of Vision*, *6*(9), 11.
- Song, J. H., & Nakayama, K. (2009). Hidden cognitive states revealed in choice reaching tasks.

- Trends in Cognitive Sciences*, 13(8), 360–366. doi: 10.1016/j.tics.2009.04.009
- Spector, B. (2007a). Aspects of the pragmatics of plural morphology: On higher-order implicatures. *Presuppositions and implicatures in compositional semantics*, 243–281.
- Spector, B. (2007b). Scalar implicatures: Exhaustivity and Gricean reasoning. *Questions in dynamic semantics*, 17, 225–249.
- Spivey, M. J., & Dale, R. (2006). Continuous dynamics in real-time cognition. *Current Directions in Psychological Science*, 15(5), 207–211.
- Spivey, M. J., Grosjean, M., & Knoblich, G. (2005). Continuous attraction toward phonological competitors. *Proceedings of the National Academy of Sciences of the United States of America*, 102(29), 10393–10398. doi: 10.1073/pnas.0503903102
- Stalnaker, R. (1978). *Assertion*. Cole, P. (ed.) *pragmatics (syntax & semantics 9)*. 315–332. Academic Press, New York.
- Steedman, M. (2012). *Taking scope*. MIT Press.
- Syrett, K., & Musolino, J. (2013). Collectivity, distributivity, and the interpretation of plural numerical expressions in child and adult language. *Language Acquisition*, 20(4), 259–291.
- Szabolcsi, A. (1997). Strategies for scope taking. In *Ways of scope taking* (pp. 109–154). Springer.
- Thothathiri, M., & Snedeker, J. (2008). Syntactic priming during language comprehension in three- and four-year-old children. *Journal of Memory and Language*, 58(2), 188–213.
- Tian, Y., & Breheny, R. (2016). Dynamic pragmatic view of negation processing. In P. Larrivée & C. Lee (Eds.), *Negation and polarity: Experimental perspectives* (pp. 21–43). Cham: Springer International Publishing. Retrieved from https://doi.org/10.1007/978-3-319-17464-8_2 doi: 10.1007/978-3-319-17464-8_2
- Tian, Y., Breheny, R., & Ferguson, H. J. (2010). Why we simulate negated information: A dynamic pragmatic account. *The Quarterly Journal of Experimental Psychology*, 63(12), 2305–2312.
- Tomlinson, J. M., Bailey, T. M., & Bott, L. (2013). Possibly all of that and then some: Scalar implicatures are understood in two steps. *Journal of memory and language*, 69(1), 18–35.
- Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science*, 247(4940), 301–306.
- Turnstall, S. L. (1998). *The interpretation of quantifiers: semantics and processing* (Unpublished doctoral dissertation). University of Massachusetts, Amherst.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological review*, 108(3), 550.
- Ussery, C. (1998). *Processing plural dps: Collective, cumulative, and distributive interpretations*. (University of Massachusetts, Amherst.)
- Van Rooij, R., & Schulz, K. (2004). Exhaustive interpretation of complex sentences. *Journal of logic, language and information*, 13(4), 491–519.
- Vehtari, A., Gelman, A., & Gabry, J. (2016). *Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC*. Manuscript, HIIT and Columbia University.
- Viau, J., Lidz, J., & Musolino, J. (2010). Priming of abstract logical representations in 4-year-olds. *Language Acquisition*, 17(1-2), 26–50.
- Vries, H. d. (2015). *Shifting sets, hidden atoms: the semantics of distributivity, plurality and animacy* (Unpublished doctoral dissertation). Utrecht University.
- Wason, P. C. (1965). The contexts of plausible denial. *Journal of verbal learning and verbal behavior*, 4(1), 7–11.
- Wason, P. C., & Johnson-Laird, P. N. (1972). *Psychology of reasoning: Structure and content* (Vol. 86). Harvard University Press.
- Winter, Y. (2001). *Flexibility principles in boolean semantics*. Cambridge, MA: MIT Press.
- Winter, Y., & Scha, R. (2015). Plurals. *Handbook of Contemporary Semantic Theory*, 77–113.
- Wojnowicz, M., Ferguson, M. J., Spivey, M., Wojnowicz, M. T., Ferguson, M. J., Dale, R., & Spivey, M. J. (2009). The Self-Organization of Explicit Attitudes. *Journal of Experimental Psychology*, 20(July 2017), 1428–1435. doi: 10.1111/j.1467-9280.2009.02448.x

- Xiao, K., & Yamauchi, T. (2014). Semantic priming revealed by mouse movement trajectories. *Consciousness and cognition*, 27, 42–52.
- Xiao, K., & Yamauchi, T. (2017). The role of attention in subliminal semantic processing: A mouse tracking study. *PloS one*, 12(6), e0178740.
- Zipf, G. K. (1949). Human behaviour and the principle of least-effort. cambridge ma edn. *Reading: Addison-Wesley*.
- Zweig, E. (2009). Number-neutral bare plurals and the multiplicity implicature. *Linguistics and philosophy*, 32(4), 353–407.

Resumé

Les ambiguïtés de phrases sont au cœur de la recherche sur la compréhension du langage depuis un certain temps. Pour les sémanticiens, ces ambiguïtés ont été utilisées pour suggérer l'existence de différents mécanismes abstraits qui pourraient s'appliquer à une même structure syntaxique au stade de l'interprétation. Pour les psycholinguistes, les ambiguïtés sémantiques ont offert un outil d'étude de la dynamique du traitement de phrases: puisque les ambiguïtés tendent à être résolues incrémentalement (c'est-à-dire avant la fin de la phrase), le schéma de traitement des phrases ambiguës peut permettre d'identifier les facteurs linguistiques et non linguistiques qui jouent un rôle dans la compréhension online.

Cette dissertation traite des théories de la compréhension du langage en explorant deux questions complémentaires: (1) comment différents sens peuvent-ils être associés à une seule tournure de phrase, et (2) comment sommes-nous capables d'accéder à ces interprétations alternatives et de les traiter pendant l'analyse syntaxique.

Pour répondre à ces questions, la présente étude se focalise principalement sur ce qu'on appelle les ambiguïtés de pluriel, qui surgissent par l'interaction entre certains prédicats et leurs arguments pluriels. Par exemple, la phrase *Amir et Milica ont construit un château de sable* a une interprétation non distributive, collective (c'est-à-dire qu'Amir et Milica ont construit ensemble un château de sable) mais aussi une interprétation distributive (c'est-à-dire qu'Amir et Milica ont chacun construit un château de sable). La plupart des approches linguistiques partent du principe que les lectures distributives dérivent d'interprétations non distributives, plus élémentaires, par l'application d'un opérateur de "distributivité" phonologiquement nul (Link, 1983; Champollion, 2014).

La première partie de cette dissertation présente deux études qui visent à identifier les mécanismes abstraits qui sous-tendent le contraste distributif/non distributif à travers un paradigme d'amorçage. Cette méthode d'amorçage est ensuite étendue à d'autres phénomènes sémantiques (c'est-à-dire des ambiguïtés de portée) dans la deuxième partie de la dissertation, dans laquelle des interactions entre pluralité et phénomène de portée sont aussi testées expérimentalement.

Pour évaluer la dynamique de la résolution des ambiguïtés, la troisième partie de cette dissertation présente une étude de suivi des mouvements de souris, conçue pour établir les caractéristiques de trajectoires de souris qui se corrént avec la prise de décision et la désambiguïsation. La méthodologie développée dans cette étude est ensuite utilisée pour analyser des données préliminaires relatives au

Mots clé

ambiguïté de phrase, sémantique, pluriels, portée, distributivité, psycholinguistique

Abstract

Sentence ambiguities have been at the center of the research on language comprehension for some time. For semanticists, these ambiguities have been taken to suggest the existence of different abstract mechanisms that may apply to the same syntactic structure at the interpretation stage. For psycholinguists, semantic ambiguities have provided a tool to analyze the dynamics of sentence parsing: since ambiguities tend to be solved incrementally (i.e. before the end of the sentence), the processing pattern of ambiguous sentences might allow identifying the linguistic and non-linguistic factors that play a role during online comprehension. This dissertation informs theories of language comprehension by exploring two complementary questions: (1) how are different meanings associated to a single sentence form, and (2) how are we able to access and compute these alternative interpretations during parsing. To address these questions, the present work mainly focuses on the so-called plural ambiguities, which arise by the interaction between certain predicates and their plural arguments. For instance, the sentence *Amir and Milica built a sandcastle* has a non-distributive, collective, interpretation (i.e. Amir and Milica together built a sandcastle) as well as a distributive one (i.e. Amir and Milica each built a sandcastle). Most linguistic approaches assume that distributive readings are derived from more basic non-distributive interpretations by the application of a covert "distributivity" operator (Link, 1983; Champollion, 2014). The first part of this dissertation presents two studies that aim to identify the abstract mechanisms underlying the distributive/non-distributive contrast through a priming paradigm. This priming method is then extended to other semantic phenomena (i.e. scope ambiguities) in the second part of the dissertation, where some interactions between plurality and scope phenomena are also tested experimentally. To assess the dynamics of ambiguity resolution, the third part of this work presents a mouse-tracking study designed to establish the features of mouse-trajectories that correlate with decision making and disambiguation. The methodology developed in this study is then used to analyse preliminary data on the processing of plural ambiguous sentences.

Keywords

sentence ambiguity, semantics, plurals, scope, distributivity, psycholinguistics