Steps Forward in the Pragmatic Understanding of Linguistic Negation

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von

Francesca Capuano

aus Maddaloni/Italien

Tübingen

2022

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation:	14.02.2023
Dekan:	Prof. Dr. Thilo Stehle
1. Berichterstatter:	Prof. Dr. Barbara Kaup
2. Berichterstatter:	Prof. Dr. Claudia Maienborn
3. Berichterstatter:	Prof. Dr. Pia Knoeferle

Kurzfassung

Negation hat seit jeher das Interesse von WissenschaftlerInnen aus unterschiedlichsten Bereichen der Forschung auf sich gezogen. Dabei fasziniert das Thema besonders deshalb, weil der logische Operator der Negation deutlich von seinem sprachlichen Gegenstück abweicht. Die sprachliche Negation wurde schon früh als markiert gegenüber der sprachlichen Affirmation angesehen. In der Tat gibt es viele Belege dafür, dass Negation mit erhöhten Verarbeitungsschwierigkeiten einhergeht. Diese Verarbeitungsschwierigkeiten können teilweise abgemildert oder ganz behoben werden - zum Beispiel durch pragmatische Lizensierung. In welchen Kontexten genau die Negation schwer zu verarbeiten ist und warum, ist bislang jedoch nicht abschließend geklärt. Die vorliegende Dissertation soll einen Beitrag zum Verständnis der sprachlichen Negation leisten, wobei insbesondere pragmatische Aspekte im Vordergrund stehen. Der Fokus liegt auf der Rolle von Alternativen und der Frage danach, wann Negationen produziert werden. Hierzu wurden drei Serien von experimentellen Studien durchgeführt. Eine typische Verwendung von Negation besteht darin, einen bestimmten Sachverhalt zurückzuweisen. Die Annahme liegt nahe, dass damit die Aufmerksamkeit auf einen alternativen Sachverhalt gelenkt werden kann, nämlich auf den Sachverhalt, der stattdessen gilt. In der ersten Studie wurde mithilfe einer Satz-Vervollständigungsaufgabe genauer untersucht, wie die durch Negation aktivierten Alternativen geartet sind. Speziell ging es um die Hy-

pothese, dass alternative Sachverhalte den verneinten Sachverhalten sehr ähnlich sind. Tatsächlich ließ sich in dieser Studie zeigen, dass die Ähnlichkeit zwischen einer negierten und deren alternativen Entität höher ist als die Ähnlichkeit zwischen zwei Entitäten, die in einem gemeinsamen affirmativen Kontext erwähnt werden (*Sie sieht keine Ziege, aber sie sieht ...* vs. *Sie sieht eine Ziege und sie sieht ...*). In der zweiten Studie wurde der Frage nachgegangen, ob Negation automatisch eine Suche nach plausiblen Alternativen auslöst. Es zeigt sich, dass plausible Alternativen auch dann automatisch aktiviert werden, wenn die Wahl der Alternativen nicht auf eine Option beschränkt ist. Beim Vergleich der Aktivierungsniveaus plausibler Alternativen (z.B. *Birne*) und semantisch assoziierter, aber unplausibler Alternativen (z.B. *Samen*) nach affirmativen und negativen Sätzen (z.B. *Das ist ein/kein Apfel*), scheint die Negation plausible Alternativen relativ stärker zu aktivieren als die Affirmation. In der dritten Studie wurde untersucht, wann Personen Negationen produzieren. Hier lag der Fokus auf der Frage, ob die Negationsproduktion durch Überlegungen zur Sprachökonomie moduliert werden kann. Speziell wurde die Hypothese untersucht, ob Negation insbesondere dann produziert wird, wenn entsprechende affirmative Aussagen besonders aufwändig sind. In einer Serie von Experimenten bekamen Probanden in jedem Durchgang zwei Kreise mit unterschiedlichen Mustern präsentiert und sollten mithilfe eines referentiellen Ausdrucks auf den Kreis referieren, auf den ein Pfeil zeigte. Manche der Muster waren schwer zu benennen. Es zeigte sich, dass Negation in der Tat umso häufiger produziert wurde, je schwieriger die Referenz mithilfe einer Affirmation bewerkstelligt werden konnte (*der Kreis ohne Punkte* vs. *der Kreis, der wie zerbrochenes Glas aussieht*). Insgesamt passen die Ergebnisse der drei Studien zu der Auffassung, dass eine wichtige kommunikative Funktion von Negation darin besteht, falsche Annahmen zu korrigieren und dadurch die Aufmerksamkeit auf plausible Alternativen zu lenken. Die Produktion von Negation wird aber offensichtlich

auch durch ökonomische Erwägung moduliert. Es bleiben viele Fragen zur Negationsverarbeitung offen, die Ergebnisse der vorliegenden Dissertation zeigen jedoch, dass eine Untersuchung der Negationsverarbeitung unter einer pragmatischen Perspektive besonders vielversprechend ist.

Abstract

Over the centuries, negation has attracted the interest of scholars from various fields. Negation aroused curiosity because of the divergence of the logical operator from its linguistic counterpart. Linguistic negation was soon pinned with a 'marked', 'inferior' status with respect to affirmation. In fact, psychological research confirmed that negation is associated to increased processing difficulties. These can in some occasions be alleviated if not eliminated altogether. Which contexts exactly make negation more or less hard to process and why, is still unclear, but the findings begin to be understood in light of a pragmatic view of negation. The current dissertation is aimed at contributing to our pragmatic understanding of linguistic negation and filling some of the gaps in the literature. Particularly, we focus on the role of alternatives and the question of when negations are produced. A typical use of negation is to reject a state of affairs. In so doing, negation might shift the attention to a different scenario (i.e. an alternative), namely the state of affairs that applies instead. Three studies were run. In the first study, we investigated the nature of plausible alternatives to negated entities through a series of cloze tasks. Specifically, we tested the hypothesis that alternatives to negation are peculiarly similar to the negated entity. Indeed, we show that the similarity of alternative entities in a negated context (e.g. *She sees no goat, but she sees* ...) exceeds that of alternatives in an affirmative context (e.g. *She sees a goat and she sees*...). In the second study, we investigated

whether negation automatically activates a search for plausible alternatives. It is evinced that plausible alternatives can be activated automatically even when the choice of alternatives is not confined to one option, and especially if the alternative is very prominent. Thus, when comparing activation levels of plausible alternatives (e.g. *pear*) and semantically associated but implausible alternatives (e.g. *seed*) after affirmative and negative sentences (e.g. *This is an apple* vs *This is not an apple*), negation appears to activate plausible alternatives relatively more than affirmation. The third study investigated when people produce negation. Here, the focus was on whether negation production can be modulated by considerations of economy of effort when compared to a concurrent affirmation. Specifically, we investigated whether negation is produced more often when concurrent affirmative statements are particularly elaborate. In a series of experiments, subjects were presented with pairs of circles filled with different patterns and were asked to refer to the circle pointed to by an arrow using a referential expression. Some of the patterns were difficult to name. In fact, the more difficult it was to refer to the relevant circle using an affirmation, the more often negation was produced (e.g. *the circle without stripes* vs. *the circle that looks like shattered glass*). Overall, the results of the three studies fit the view that an important communicative function of negation is to correct false assumptions and thereby draw attention to plausible alternatives. However, the production of negation is also modulated by economic considerations. A lot remains to be understood about how negation is processed, but the current results confirm that an investigation of negation processing from a pragmatic perspective seems particularly promising.

Contents

Chapter 1

Introduction

Negation is a universal property of natural language, claimed to be absent from any other animal communication system (Horn, 1989): despite the variations, every human language has devices to achieve truth value reversal (Dahl, 1979; Horn, 1989; Miestamo, 2007).

For centuries, negation has been object of study for logicians, philosophers, linguists and psychologists. Its complex behaviour and various expressions in natural language clash with the straightforward functioning of negation as a logical operator. In fact, ever since Plato and Aristotele, natural language negation was pointed at as asymmetrical - somewhat inferior - with respect to affirmation. The asymmetricalists noticed how negative utterances tend to be less specific and less informative than affirmatives. Negation is - more often than not - the marked member of the pair (De Swart, 2009), and it arguably presupposes affirmation.

Psychological studies traditionally confirmed this asymmetry: while negation seems to be ubiquitous, even more so in spoken language (Tottie, 1982), it is often associated with processing and integration difficulties (e.g. Deutsch et al., 2006; Dudschig et al., 2019;

Fischler et al., 1983) that cannot even be eased by additional preparatory time (Dudschig & Kaup, 2020a). This is true not only for linguistic, but also for pictorial negation (Dudschig & Kaup, 2021). Over the years, the literature has gained some insight into the nature of this asymmetry and on the peculiarities of conversational negation, especially thanks to a growing interaction of the linguistics semantic and pragmatic traditions with psychology and varied experimental paradigms (for a recent overview, see Dudschig, Kaup, Liu, et al., 2021).

1.1 The Disadvantage of Negation

A long line of research from the 60s and 70s indicated that negative statements are a lot harder to process than affirmative statements (e.g. Carpenter & Just, 1975; Clark & Chase, 1972, 1974; Just & Carpenter, 1971; Wason, 1961; for an overview, see Kaup & Dudschig, 2020). The majority of these studies timed truth value judgments of negative vs. affirmative sentences with respect to world knowledge or a given visual world (e.g. *The star [is]/[is not] above the plus* coupled with a picture of a star standing above/below a plus sign in Clark & Chase, 1972). The consistent finding across experiments was an increase in response times to negative sentences, that was unjustifiable on the sole basis of additional reading material (e.g. *not*).

Two different patterns emerged relative to the effect of truth value. Some studies resulted in a main effect of truth value, with false sentences taking longer to verify independently of polarity (e.g. Eifermann, 1961; Just & Carpenter, 1971; Wason, 1961). In other cases, an interaction of truth value and polarity arose, with true affirmatives being easier than false affirmatives, but true negatives harder than false negatives (Carpenter & Just, 1975; Clark & Chase, 1972, 1974; for a recent example, see Rück et al., 2021).

The results were explained through the use of different strategies to 'get rid' of negation. According to the *true model*, subjects would represent both pictures and sentences as embedded propositions, and then compare them constituent by constituent, while keeping track of incongruencies. According to the *conversion model*, subjects would try and convert negative to affirmative propositions (e.g. *5 is not even* to *5 is odd*) before verifying the truth value of the sentence (Chase and Clark, 1972; Clark, 1976). The first model explains the first pattern of results (main effect of truth value), while the second model accounts for the other pattern (truth value x polarity interaction). Mayo et al. (2004) found evidence suggesting that the two different encodings might co-exist, and the preference for the one over the other might depend on the availability of the alternative schema: bi-polar descriptions (e.g. *tidy/messy*) are more likely to be *converted* than uni-polar descriptions (e.g. *adventurous*).

Whichever the preference for one over the other strategy depending on the context, these accounts fail to explain just why negation processing should involve the adoption of specific strategies to either reduce it or to get rid of it. In other words, why is negation so dispreferred and harder to process with respect to affirmation?

It was suggested that the disadvantage is due to increased transformational complexity, which requires an additional step to reach the deep structure when compared to affirmation (Chomsky, 1957). Subsequent studies did not provide decisive evidence in favour of the psychological reality of transformational grammar (e.g. Engelkamp & Hörmann, 1974; Fillenbaum, 1966; Hörmann, 1971). A second explanation was given in terms of a general inhibiting effect due to early associations of negation with privation and prohibition during childhood. Some evidence was adduced to this idea: negators typically associated with prohibition result in more processing difficulties than either negators exclusively associated with denial (Eifermann, 1961; but see Wales & Grieve, 1969) or

senseless syllables replacing negators (Wason & Jones, 1963); moreover, explicit negation seems harder to process than implicit negation (Clark, 1976, S. Jones, 1968). Finally, a widely entertained explanation throughout this discussion involved pragmatic differences between affirmative and negative sentences. A large body of research showed that the difficulty associated with the processing of negation could be modulated by varying the context of presentation of the sentences.

1.2 The Pragmatic Explanation

A plausible explanation for these processing difficulties lies in the pragmatic differences between affirmative and negative sentences. Although the idea was not new, Wason (1965) brought to the attention of the psycholinguistics community that the ease of negation processing might depend on appropriateness of use in the context. Since most of the literature from the 60/70s presented affirmative and negative sentences out of a negation-licensing context, this resulted in the negative sentences being less pragmatically felicitous than their affirmative counterparts.

Considerable evidence was adduced to the idea that the ease of processing negation depends on its pragmatic felicity. One factor affecting the pragmatic felicity of negation is the concept of exceptionality: Wason (1965) found completions to negative sentences to be facilitated when an exceptional item is described in terms of a missing property with respect to a residual class. For example, the sentence *Circle N.4 is not...* is easier to complete with an appropriate attribute (e.g. *red*) when it refers to a blue circle amidst seven red circles, but this only holds if the seven red circles are coded in terms of a subsidiary class with respect to the class of interest (e.g. one circle is blue and the rest are red) as opposed to when the two classes are perceived as existing independently from one

another (e.g. seven circles are blue and one is red). Similarly, De Villiers and Flusberg (1975) indicated that children are faster at completing negative sentences that refer to an exception. Additionally, with age, they become more sensitive to confusability: completions are even faster if the exception is easily confusable with the contrast class, as in *This is not a horse* referring to a cow among several horses vs. *This is not a flower* referring to a shoe among several flowers. Further evidence is provided by Cornish (1971). In this study, subjects were presented with circles that were filled with two colours in varying proportions (e.g. red and blue). The higher the proportion of red in the circle, the easier could the subjects understand and complete a sentence like *The circle is not all red*. Glenberg et al. (1999) showed a further way to use negation in a pragmatically felicitous manner by introducing the negated attribute dimension in the preceding context, thereby facilitating negation processing. Lüdtke and Kaup (2006) similarly demonstrated that negation is easier to process when the negated proposition is explicitly mentioned or at least constitutes a plausible assumption in the preceding context. Indeed, Albu et al. (2021) find that negative sentences are facilitated in contexts of plausible denial, but the use of a discourse marker like *contrary to expectations* is not enough to set the facilitation into motion: the expectation to counter needs to be already introduced (more or less explicitly) in the previous context. Finally, N400 patterns align to their usual behaviour in truth value judgment tasks when negation is used to reject a plausible misconception (Nieuwland & Kuperberg, 2008; cf. Dudschig et al., 2019; Fischler et al., 1983).

In conclusion, negation seems easier to process in contexts licensing its use via pragmatic principles. Crucially, these findings are all in line with the general idea that negation is often used to indicate deviations from expectations and/or to correct a false presupposition (Clark & Clark, 1977; Givón, 1978; Glenberg et al., 1999; Horn, 1989; Wason, 1965, 1972) that, if not necessarily held true by the listener, is at least relevant given the

context (for a different view, see Giora, 2006). How did a deviation from expectations/-correction of a false presupposition come to be among the most prevalent communicative aims of negation? The next paragraph will introduce a possible explanation for this presuppositional implicature associated to negation.

1.3 The Origins of the Presuppositional Implicature

According to Horn (1989), the pragmatic asymmetry between negation and affirmation might originate from the interaction and mutual constraint between the two forces of Speaker's Economy and Auditor's Economy. Speaker's Economy would, if unrestrained, tend towards the reduction of the lexicon to one word denoting every possible meaning, whereas Auditor's Economy would push towards a vocabulary of as many words as there are meanings. The two forces constrain each other by satisfying the speaker's need to convey the relevant information while minimizing his efforts.

In the Neo-Gricean account of nonlogical inference, the two forces correspond to the Q (Quantity, hearer-oriented) and R (Relation, speaker-oriented) principles (Horn, 1984). The Q principle generates upper-bounding implicata, producing the inference that the speaker did not employ a stronger, more informative form because they were not in the position to do so. The R principle generates lower-bouding implicata, with the inference that the speaker intended to convey the meaning of a stronger form, but could not employ it due to (usually) social or cultural constraints.

Within the framework of the Neo-Gricean analysis of conversational implicature, accounts have been proposed for the tendency of negation to be used in a context that involves the expectation or plausibility of the negated state of affairs. The supposition might arise from the interaction of the Q principle with the principle of negative uninfor-

mativeness. As negation is usually less informative than affirmation, one must assume a context where it is just as informative as required (and no more than that). Those are precisely the contexts where the truth of the affirmative state of affairs is being considered. For example, a sentence like *Bogota isn't the capital of Peru* is felicitously uttered in a context where the interlocutor believed the positive supposition to be possibly the case, but in a context where the interlocutor was additionally interested in knowing what the capital of Peru is, the sentence gives rise to the inference that the speaker does not possess that information. Only in a context where this information is relevant and the speaker can be assumed to possess that knowledge, the negative sentence is infelicitous because unnecessarily weak.

This prototypical usage of negation, which involves the presence of the presupposition of the negated state of affairs, might 'infect' even cases where affirmation and negation are equally informative. For example, these expectations seem to spill over to the use of contradictory predicates (e.g. *Twenty-four is not an odd number* vs. *Twenty-four is an even number* in Wason, 1961). Even though, from a logical perspective, the two sentences should be equivalent as they convey membership to one of two mutually exclusive classes, a sentence like *Twenty-four is not an odd number* presented in isolation might be harder to process than the affirmative *Twenty-four is an even number* simply because it does not fulfill the pragmatic functions *normally* attributed to negation.

The asymmetry between affirmatives and negatives therefore seems to lie at the pragmatic level. Comprehenders, in turn, seem to be sensitive to these pragmatic aspects, processing negation faster if it is used in line with these pragmatic principles. As far as we know, Nordmeyer and Frank (2015) is the only study that established a direct link between negation production and comprehension, while explicitly relating the felicity of negation to general pragmatic principles. The authors showed that negation is

spontaneously produced more often as it becomes more informative and more relevant. Informative utterances should reduce uncertainty in the state of affairs: we can think of informativity as the update of prior beliefs. Relevant utterances address a question that is pertinent in the current context: in fact, relevance is often formalised by the concept of *Question Under Discussion* (QUD, see Roberts, 1996/2012). In each trial of their experiment, Nordmeyer and Frank presented subjects with a depiction of four characters. The trials differed in the proportion of characters holding the same, two recognizable objects (e.g. 0/4 or 1/4 or 2/4 or 3/4 or all characters were holding two apples). Apart from whether they were holding the objects, the characters presented within a trial were identical. Subjects were asked to complete sentences of the form *[NAME] has...* in reference to one of the four characters, which was highlighted by a red rectangle. The percentage of negative completions grew, on the one hand, as the use of negation increased the probability of identifying the correct referent, therefore reducing uncertainty (i.e. informativity); on the other hand, as it addressed a *relevant* underlying question (e.g. *Bob has no apples* only becomes relevant when apples are present at all). In addition, the same depictions were paired with sentences of the form *[NAME] [has]/[has no] [ITEM]* and presented to a different batch of participants for a truth value judgment task to investigate comprehension ease of these items. Crucially, reaction times to true affirmative and negative sentences were predicted by their surprisal, calculated from the production probabilities derived from the speakers' task. The finding strongly supports a pragmatic view on the nature of the processing difficulties of negation. At the same time, informativity is highlighted as a drive for the production of negation. As noted by Nordmeyer and Frank (2015), these results support a Gricean perspective on negation processing (Grice, 1975): speakers follow – and listeners expect them to follow - the conversational maxims of informativity and relevance.

1.4 Negation and Alternatives

Given that the prototypical usage of negation seems to be that of rejecting or correcting a presupposition, it remains to be understood whether it additionally gives rise to the inference of an alternative, 'correct' scenario.

Alternatives are key to the successful comprehension of certain types of utterances. Following the work of Rooth (1992) on alternative semantics, focused items evoke alternatives which are relevant to the interpretation of an utterance (e.g. *[Mary] likes Sue* evokes the set of propositions of the form *x likes Sue*, whereas *Mary likes [Sue]* evokes the set of propositions of the form *Mary likes x*). Recent priming studies have tapped into the psychological reality of this semantic theory, showing that listeners do infer alternatives to focused items, even when these are not explicitly mentioned in the previous context (Braun & Tagliapietra, 2010; Husband & Ferreira, 2016; Yan & Calhoun, 2019). Using cross-modal lexical decision tasks, Braun and Tagliapietra (2010) investigated semantic priming of sentences differing in intonation pattern and in the semantic relation between the final word and the target word in Dutch. When an utterance was contrastively accented (e.g. *In **Florida** he photographed a **flamingo***), contextual alternatives (e.g. *pelican*) were activated preferentially compared to a neutral intonation condition, but non-contrastive semantic associates (e.g. *pink*) were not. Husband and Ferreira (2016) explored the time course of the priming effect in English, finding evidence for an initial activation of both contrastive and non-contrastive associates, followed by the selective deactivation of non-contrastive associates in the case of contrastively accented sentences. Yan and Calhoun (2019) shows preferential priming of contextual alternatives in Mandarin Chinese when the focus is realised via prosodic prominence, but not when it is realised only syntactically.

Alternatives might be equally relevant to the comprehension of negation. From a

logical perspective, negation translates to the complement set of its scope (any element belonging to it is an equally plausible alternative). Psychologically though, only a limited number of these elements can be considered as plausible alternatives. In Kruszewski et al. (2016) speakers judged a sentence like *It's not a dog, it's a wolf* as more plausible than a sentence like *It's not a dog, it's a screwdriver*. Seen from the perspective of set theory, any member of the complement set of *dog* should be consistent with *not a dog*. In fact though, it seems more natural to assume that someone might mistake a wolf for a dog, rather than a wolf for a screwdriver. Alternativehood implies some level of exchangeability that is contextually defined, therefore plausible alternatives are context-dependent. But whereas any member of the complement set of an entity could potentially be a plausible alternative given enough contextual pressure, some entities tend to be 'better alternatives' than others, in that they are relevant across a large number of contexts. In this sense, the notion of alternativehood is tied to that of contextual similarity: as good alternatives are interchangeable with the negated entity, they can be expected to occur in similar contexts. Starting from these considerations, Kruszewski et al. (2016) show that the plausibility of alternatives introduced by negation can be effectively modeled in terms of contextual similarity with respect to the negated entity through Distributional Semantic (DS) representations. Distributional Semantic Models (DSMs) are computational models deriving vectorial distributed representations of meaning out of information extracted from corpora (Landauer & Dumais, 1997; Lenci, 2018; Lund & Burgess, 1996). The intuition behind the models is that words that occur in similar contexts will tend to have similar meanings (Harris, 1954). For example, both *sea* and *ocean* will often be used in the same sentence as *water*, *fish*, or *sail*. For each word, the models return a high-dimensional vector that encodes information on which contexts the word occurs in. Words that occur in similar contexts will result in similar vectors. A cosine similar-

ity score can be calculated between pairs of vectors to reflect the degree of similarity in meaning between the two corresponding words. The models effectively construct a 'semantic space' where the distance between words reflects their semantic distance. Even though with limitations and criticisms, DSMs are able to reflect cognitive similarity measures such as those derived from priming studies or similarity judgments (Bruni et al., 2014; Günther et al., 2016; Günther et al., 2019; M. N. Jones et al., 2006; Lenci, 2008). Kruszewski et al. (2016) show a correlation to exist between the distributional semantic similarity of the negated entity (*dog*) and its alternatives (*wolf* vs. *screwdriver*), and the plausibility judgments on the sentences: the more similar the two entities (e.g. *dog* and *wolf*), the more plausible was the sentence rated. Possibly, higher similarity corresponds to higher confusability between the two entities, therefore licensing the assumption of the speaker on a false presupposition. Rather than denoting the whole complement set indiscriminately, natural language negation therefore seems to act as a graded similarity function, defining a probability distribution over a restricted set of alternative scenarios. It is unclear though whether listeners activate likely alternatives to negation.

Evidence suggests that – whether mediated by the representation of the negated state of affairs (*two-step simulation hypothesis*, see Kaup et al., 2005; Kaup et al., 2007; Kaup & Zwaan, 2003) or not (Anderson et al., 2010; Giora, 2006; Orenes, 2021; Tian et al., 2010; Tian et al., 2016) - speakers eventually reach a representation of the alternative state of affairs as a result of processing negative statements. Kaup et al. (2006) had their subjects read sentences involving contradictory predicates (e.g. *The door is (not) closed/open.*). The subjects were then presented with a picture and asked to name aloud the entity depicted. The entity was varied according to whether it corresponded to the described or to the negated state of affairs. According to the sentence *The door is closed*, a closed door would correspond to the actual state and an open door would not. On

the other hand, according to the sentence *The door is not closed*, an open door would correspond to the actual state and a closed door would not. In order to investigate the temporal characteristics of the representational process, the delay at which the image was presented was varied. One half of the subjects was presented with the image after 750 ms, the other half only after 1500 ms. With a delay of 750 ms, an effect of agreement was found in relation to the actual state in affirmative sentences (e.g. a closed door after the sentence *The door is closed.*), but not in negative sentences (e.g. an open door after the sentence *The door is not closed.*). With a delay of 1500 ms, this effect was also present for negative sentences - the subjects had thus mentally represented the actual state at this point in time, while at an earlier point in time the state to be negated was represented. In cases like this though, the person addressed can easily infer the alternative state (an open door). In fact, using the example of the door, *open* and *closed* represent the only two possible states. The actual state of affairs (the alternative) is therefore confined to one possibility.

1.5 Aims of this Dissertation

The aim of this dissertation is to further explore the extent of the pragmatic peculiarities of conversational negation and the asymmetry between negation and affirmation. In particular, we investigate the issue of the alternative scenarios that can be inferred from negation, and whether they follow along with the pragmatic peculiarities attributed to negation (correction, rejection of a plausible assumption).

In Study 1, we will investigate whether alternative scenarios that are inferred from negation actually display peculiarities attributable to the pragmatics of negation. Kruszewski et al. (2016) did not compare negated and affirmative sentences. It is therefore not infor-

mative for the assumption that negation specifically comes with the presentation of very similar alternatives: the possibility persists, that affirmative sentences elicit even higher similarities. In fact, the correlation between the similarity of nouns within a sentence and plausibility ratings might hold true in general (e.g., people might also prefer an affirmative sentence such as *This is a dog and that is a wolf* over *This is a dog and that is a screwdriver*): without a baseline comparison condition we will not be able to tell whether negation specifically prompts very similar alternatives beyond what can be suggested by for example lexical association and contiguity. It is therefore necessary to investigate the question of whether specifically negation results in the production of similar scenarios, by comparing the results to an affirmative conjuncting context.

In Study 2, we will explore whether plausible alternatives are inferred during online negation processing. In Study 1 subjects were explicitly prompted for an alternative. To our knowledge, it has not been investigated whether negation automatically activates a search for plausible alternatives in case no alternative were explicitly solicited. Under those circumstances, we might expect plausible alternatives to be activated preferentially after negative rather than after affirmative statements. This would be in line with existing evidence on activation levels after negation processing. MacDonald and Just (1989) showed that negation decreases the activation level of the mental representation of the negated entity. On the other hand, concepts related to the negated concept were not significantly less active than in the case of affirmation. Therefore, negation inhibited the activation of the negated concept but not of its associates. Crucially, no differentiation was made between associates that are also plausible alternatives with respect to the negated entity, and associates that are not. Therefore, their results do not preclude the possibility of a different pattern for plausible vs. implausible alternatives.

In Study 3, we will preliminarily investigate whether speaker's economy can mod-

ulate the production of negation to refer to an alternative in case no presupposition is involved. Most work on negation focused on the comprehension, as well as the completion of sentences that already encode polarity: there is little evidence for when negation is spontaneously produced. Nordmeyer and Frank (2015) showed that negation is spontaneously produced more often as it becomes both more relevant and more informative (see 1.3). Watson (1979) prompted children to freely describe an entity so that it could be differentiated from a comparison entity. Negation was produced more often than affirmation to describe the referent (i.e. a white horse) when it lacked a property with respect to the contrast item (e.g. a horse with spots) but close to never when the contrast item differed along an attribute dimension (e.g. a black horse), therefore again following a principle of informativity. Relatedly, Beltrán et al. (2008) show that negation is spontaneously produced when speakers might not have enough information to grant affirmation. Practically, negation is produced more often when an affirmative description of the state of affairs is not available. This availability was modulated by the use of bipolar vs non-bipolar attributes (e.g. *big/small* vs. *red/green/blue/...*), as well as by whether the alternative state of affairs was explicitly mentioned in the context. In Experiment 1, participants were asked to complete short narratives describing the information reported on a source as erroneous (e.g. *In a magazine there was some wrong information. It talked about the [size/color] of a car. Juan realized that the information mistakenly stated that the car was [big/red]. In fact, the car...*). In Experiment 2, the alternative attribute was explicitly stated (e.g. *They discussed if the car was [big or small]/[red or green].*). More negative completions were produced in Experiment 1 in the non-bipolar condition, whereas the difference was absent in Experiment 2. Therefore, the effect seems attributable to the availability of the state of affairs rather than the attribute type. Producing a negation avoids the violation of the maxim of quality: the subjects lack adequate evidence for the

actual state of affairs. Taken together, the reported studies suggest that negation production seems to align to general pragmatic principles. In the present study, we investigate whether economy also plays a role. In other words, speakers might be induced to produce negation not only on the basis of informativity and relevance expectations, but also following a general principle of least effort (Zipf, 1949).

Chapter 2

Results

2.1 Declaration on the Contribution of Others

This dissertation largely draws on three co-authored manuscripts. This is most clearly recognizable in Chapter 2, whereby Section 2.2 presents the results of Capuano et al. (2021) (contributions acknowledged in Table 2.1), Section 2.3 presents the results of Capuano et al. (2023) (contributions acknowledged in Table 2.2) and Section 2.4 presents the results of Capuano et al. (2022) (contributions acknowledged in Table 2.3).

Table 2.1

Contributions of Capuano et al. (2021).

Author	**Author position**	**Scientific ideas %**	**Data generation %**	**Analysis & interpretation %**	**Paper writing %**
Francesca Capuano	1	40	100	70	70
Carolin Dudschig	2	30	0	10	10
Fritz Günther	3	0	0	10	10
Barbara Kaup	4	30	0	10	10
Title of the paper:		Semantic Similarity of Alternatives Fostered by Conversational Negation			
Status in publication process:		published			

Table 2.2

Contributions of Capuano et al. (2023).

Author	**Author position**	**Scientific ideas %**	**Data generation %**	**Analysis & interpretation %**	**Paper writing %**
Francesca Capuano	1	90	50	60	60
Theresa Sorg	2	0	50	20	30
Barbara Kaup	3	10	0	20	10
Title of the paper:		Activation Levels of Plausible Alternatives in Conversational Negation			
Status in publication process:		published			

Table 2.3

Contributions of Capuano et al. (2022).

Author	**Author position**	**Scientific ideas %**	**Data generation %**	**Analysis & interpretation %**	**Paper writing %**
Francesca Capuano	1	40	100	80	80
Carolin Dudschig	2	30	0	10	10
Barbara Kaup	3	30	0	10	10
Title of the paper:		Using Circles Games to Investigate the Referential Use of Negation			
Status in publication process:		published			

2.2 Semantic Similarity of Alternatives Fostered by Conversational Negation

Starting from the evidence that a negated sentence is rated as more acceptable the higher the presented alternative scores in DSM measures of similarity (Kruszewski et al., 2016), the current study was designed to answer two key questions that remained open: (a) is this phenomenon also seen in production, in that speakers tend to produce alternatives to negation that score high in DSM similarity measures? (b) in how far is this higher similarity specific to negation if we compare it to standard language use?

Regarding the first point, if negation elicits very similar alternatives, then these are the words participants should be expected to produce when prompted to do so. Additionally, production tasks partly overcome the limitations of selecting item material (as opposed

to comprehension tasks): although the sentential contexts are given (e.g. *There is no X here, but there is Y*), a production task allows subjects to freely produce any alternative Y they deem appropriate without limitations on the solutions space, possibly resulting in a clearer overview of actual language usage. As for the second point, the correlation of similarity scores and plausibility judgments might not be exclusive to negation, therefore we needed to introduce comparison baselines in order to show that negation indeed evokes alternatives that are particularly similar to the negated entity. As a comparison baseline we used affirmative sentences with the *and* connector.

Subjects completed a series of cloze tasks: in each task, they were asked to complete negative sentences suggesting alternativehood with respect to a negated noun (e.g. *There is no goat here, but there is ...*) and/or affirmative conjuncting sentences that were kept structurally as similar as possible to the corresponding negative ones (e.g. *There is a goat here, and there is ... there*). As negated nouns we always used the 50 nouns employed by Kruszewski et al. (2016) (e.g. *goat*). We employed a range of minimally constraining sentential contexts across experiments (see Table 2.4). Cosine similarity scores were calculated between each negated noun/completion noun pair for negative sentences, and between each given noun/completion noun pair for affirmative sentences. All the cosine similarity scores were derived from the same semantic space employed in Kruszewski et al. (2016) with the help of the LSAfun package (Günther et al., 2015) The semantic space is the one that performed closest to empirical results throughout a series of evaluation tasks (e.g. semantic relatedness, synonym detection) in Baroni et al. (2014). We expected the affirmative baseline to already give rise to the production of very similar entities, as *and* is typically associated with coding co-occurrence of state of affairs (Mauri & Van der Auwera, 2012), and therefore can be expected to suggest contiguity. Nevertheless, as alternativehood goes beyond contiguity, additionally imply-

Table 2.4

Summary of the Results of the Main Analysis.

Experiment	Sentence Type	Task	Polarity	Sentence	*p*
1	there	1a	negative	There is no X here, but there is Y.	
		1b	affirmative	There is X here, and there is Y there.	
	this	1a	negative	This is not X, it is Y.	
		1b	affirmative	This is X, and that is Y.	0.15
2	there	2	negative	There is no X here, but there is Y.	
			affirmative	There is X here, and there is Y there.	<0.001
3	this	3	negative	This is not X, it is Y.	
			affirmative	This is X, and that is Y.	0.30
4	this	4	negative	This is not X here, but it is Y.	
			affirmative	This is X here, and that is Y there.	0.18
5	this	5	negative	This is not X here, it is Y.	
			affirmative	This is X here, and that is Y there.	0.49
6	see	6	negative	(Pron) see(s) no X, but (Pron) see(s) Y.	
			affirmative	(Pron) see(s) X and (Pron) see(s) Y.	<0.001
7	want	7	negative	(Pron) want(s) no X, but (Pron) want(s) Y.	
			affirmative	(Pron) want(s) X and (Pron) want(s) Y.	<0.001

ing the substitutability of entities, we expected even higher similarity scores for entities produced in negative sentences. Showing this would suggest that negation indeed has the specific function to prompt very similar alternatives. Responses to the cloze tasks were crowdsourced via Amazon Mechanical Turk. The survey link instructions requested the workers to be native English speakers.

In order to get a more complete overview of the distribution of semantic similarity in natural language, we additionally employed four reference points to assess whether the words produced by participants can actually be considered ‘similar’ on an absolute level, as would be expected if negation indeed involves the presentation of very similar alternatives. The reference points were the same for all experiments and were calculated with respect to these 50 nouns. The first reference point is the mean similarity score of the 50 nouns with their respective closest neighboring noun in the semantic space. A second reference point was calculated as the mean of the average cosine similarity score of each noun with their free associates, normalized by number of participants producing the free

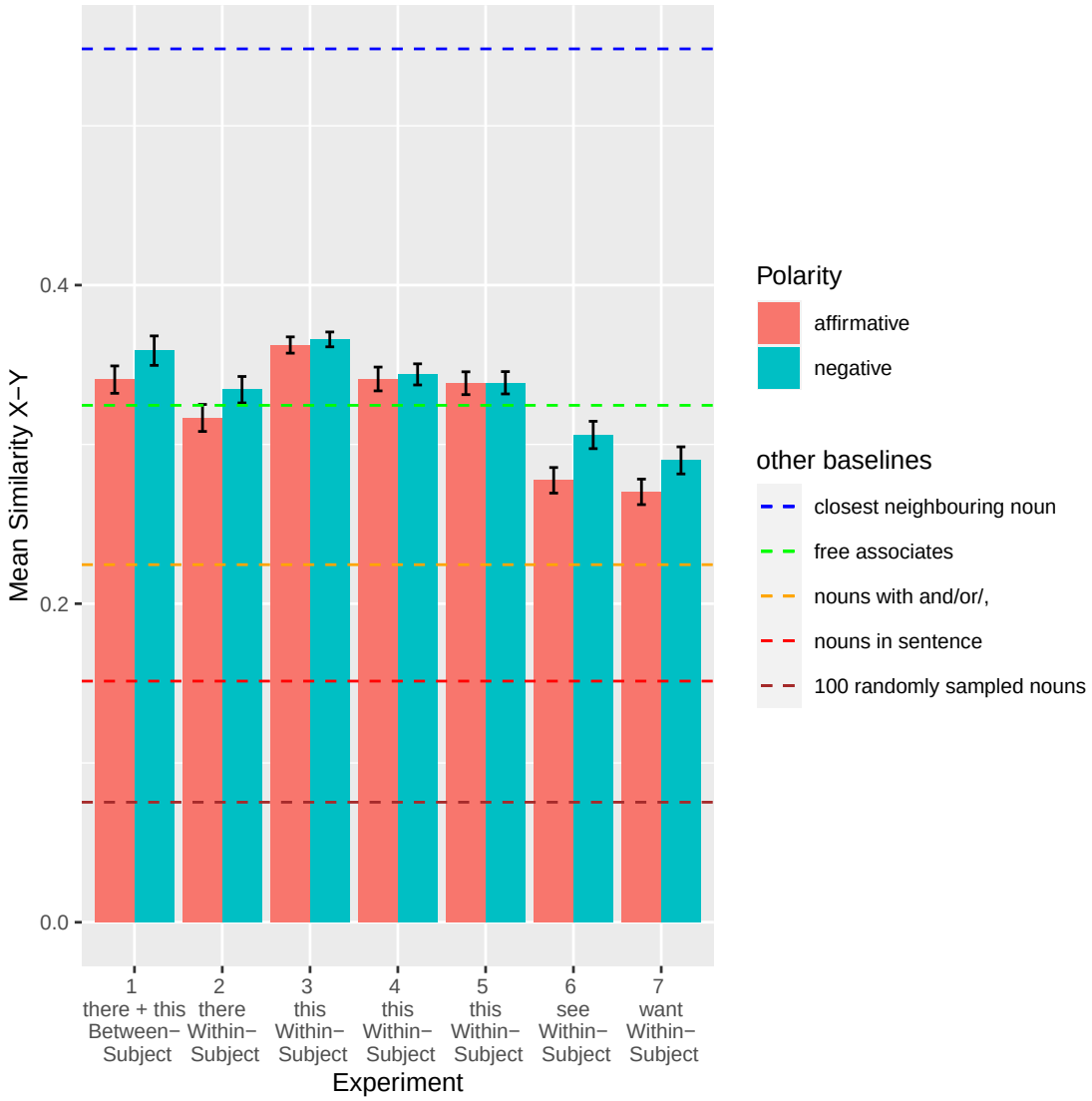
associate response. Free associates and their production frequencies were extracted from the USF free association norms (Nelson et al., 2004), which covered 42 out of the 50 nouns. The third point is the mean of the average similarity score of each noun with all the nouns it co-occurs with within the constructions [Noun] [*and*/*or*/*,*] [optional: *the*/*a*] [Noun], weighted by the co-occurrence frequencies. The fourth reference point is the mean of the average similarity score of each noun with all the nouns it co-occurs with within a sentence, weighted by the co-occurrence frequencies. Co-occurrences for the third and fourth reference points were extracted from the same corpus that was used to construct the semantic space. The last point is the mean of the average cosine similarity score of each noun with 100 randomly sampled nouns.

Overall, the average similarity scores derived from the tasks attest themselves around the level of free associates (see Figure 2.1 and Table 2.5), which seems sensitive, as the minimally constraining contexts might have rendered the tasks akin to a free association task. Free associates are the most accessible words given a cue word, where ease of retrieval seems to be affected both by associative knowledge and by aspects of meaning (Nelson et al., 2004). In fact, they are well above words co-occurring in the *and/or/,* contexts found in a corpus, which likely span a wide range of constraints.

Fundamentally, comparisons were run between negative and affirmative sentences through linear mixed effect models. For three out of four sentential contexts, similarity scores from negative sentences lie significantly above their affirmative counterparts (see Figure 2.1 and Table 2.4): on average, negation seems to indeed prompt very similar alternatives. The effects appear to be quite strong, which, together with the consistent results of Experiments 3 to 5, suggest the null effect to be specific to the *This* context. On the other hand, it is not the case that alternativehood prompting is exclusive to negation. Many of the pragmatic functions typically ascribed to negation can also be conveyed

Figure 2.1

General Baselines and Mean Cosine Similarity Scores.



Note: Mean cosine similarity scores of affirmative and negative conditions for each experiment were aggregated by subject. Error bars represent ± SE of the means.

by affirmation (Giora, 2006). Contrastive negation is not the only means of expressing contrast: similarly, contrast can be conveyed by intonational patterns, cleft constructions and word order (Silvennoinen, 2020). In the case of the *This* context, the affirmative sentence might equally convey a corrective reading, leading to the production of a substitute

Table 2.5

Comparisons of Each Polarity Condition Against the Free Associates Baseline.

Experiment	Sentence Type	Polarity	p_1	t_1	p_2	t_2
1	this + there	affirmative	0.07	1.88	0.07	1.87
		negative	<0.001	3.72	<0.001	3.82
2	there	affirmative	0.35	−0.93	0.41	−0.82
		negative	0.24	1.18	0.30	1.05
3	this	affirmative	<0.001	7.43	<0.001	3.69
		negative	<0.001	8.89	<0.001	4.02
4	this	affirmative	<0.05	2.21	0.10	1.67
		negative	<0.01	2.93	<0.05	2.45
5	this	affirmative	0.06	1.94	0.12	1.58
		negative	<0.05	2.02	0.14	1.50
6	see	affirmative	<0.001	−5.86	<0.001	−6.22
		negative	<0.05	−2.17	<0.05	−2.06
7	want	affirmative	<0.001	−6.78	<0.001	−7.18
		negative	<0.001	−4.06	<0.01	−3.41

state of affairs. On average though, this contrastive function seems to be more peculiar to negation. It is in fact widely acknowledged that one of the core pragmatic purposes of negation is that of correcting a false presupposition (Givón, 1978; Wason, 1965). A false presupposition seems more justified if there is a reason to confuse two states of affairs (presupposed vs. actual), therefore a minimal difference between the two renders the use of negation more felicitous.

In addition to the planned analysis we conducted a post-hoc analysis in order to investigate whether alternatives might be constrained on a relational level with respect to the negated entity. A stronger preference for paradigmatic relationships was expected in the case of negative sentences. Indeed the percentage of cohyponyms detected among the completions is higher for negative sentences than for affirmative sentences in 5 out of 6 experiments. The speculation is additionally supported by higher average path-length similarity scores in the negative conditions for most tasks. It is unclear though - and in

need of further clarification - whether the paradigmatic/syntagmatic distinction is actually at work or the preference can be reduced to similarity alone.

Our results confirm the idea that linguistic negation acts very differently from a logical operator, proving itself to be in fact highly restrictive in the suggestion of alternative scenarios: naturally produced alternatives are very similar to the negated entity. Furthermore, the restriction appears to be even tighter than in the case of affirmation, where entities are expected to be limited by the probability of co-occurring in the same scenario. Whereas the logical use of negation would be largely permissive in the selection of plausible alternatives, conversational negation thus acts even more restrictively than conjunction within an affirmative context. [1]

2.3 Activation Levels of Plausible Alternatives in Conversational Negation

The previous study confirmed that the preference for highly similar alternatives is indeed specific to negation, going beyond a general preference for semantically similar nouns within the same sentence. In that study subjects were explicitly prompted for an alternative. To our knowledge, it has not been investigated whether negation automatically activates a search for plausible alternatives in case no alternative were explicitly solicited.

Seen from the perspective of set theory, any member of the complement set of *dog* would be consistent with *not a dog*. As we have seen though, some entities are more likely alternatives than others: in the case of words that don't relate to a direct opposite, negation acts as a graded similarity function that produces a probability distribution over a restricted set of alternatives (Kruszewski et al., 2016). Given this uncertainty, it is

[1] Reprinted from the previously published article Capuano et al. (2021), Copyright 2021, Open Access. See Appendix A for full article.

unclear whether listeners still activate likely alternatives in their mental representation.

Alternatives have been shown to play a significant role in the comprehension of utterances involving focus. Several studies seem to suggest that subjects indeed activate alternatives to focused items as part of the comprehension of an utterance (Braun & Tagliapietra, 2010; Husband & Ferreira, 2016; Yan & Calhoun, 2019), but the same has not been investigated for alternatives to negated entities.

Based on these considerations, the present study investigates whether negation leads to the activation of plausible alternatives. For this purpose, we constructed affirmative and negative prime sentences of the form *This is [a/an]/[not a/an] X* (German *Das ist [ein/e]/[kein/e] X*). The experiments were run in German. The presentation of the sentences was followed by a lexical decision task. Targets could either constitute a plausible alternative (e.g. *pear*) or a semantically related but implausible alternative (e.g. *seed*) with respect to the entity in the prime sentence (e.g. X = *apple*). Unrelated words (e.g. *brush*) were also presented as targets for the sake of a manipulation check: as per literature, we expected both plausible alternatives and semantically related words to be responded to faster than unrelated words. Non-words were also used as targets to complete the lexical decision task. The experiment was conducted in German. If negation leads to a search for alternatives, the difference in RTs between negative and affirmative sentences is expected to be smaller in the case of plausible alternatives, compared to semantically related but implausible alternatives. More specifically, an interaction effect of Sentence Polarity and Prime-Target Relation is expected, with negation facilitating responses to Plausible Alternatives.

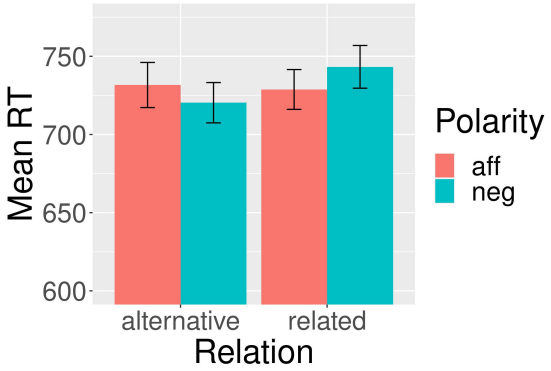
Sixty German native speakers took part in the pilot study. Forty-eight experimental items were constructed. We controlled for length and frequency between targets across target relations. The mean length for the *alternatives* was 6.10 ($SD = 1.75$), 5.81 ($SD =$

1.88) for the *related* and 5.94 ($SD = 1.37$) for the *unrelated* targets. Alternatives did not differ from related (t(47) = 0.75, p = 0.45), nor from unrelated (t(47) = .60, p = 0.55) targets. Related and unrelated targets were also paired (t(47) = -0.37, p = 0.71). Target frequencies were extracted from the deWaC corpus. The mean frequency for the alternatives was 23788 ($SD = 46004$), 24260 ($SD = 52032$) for related and 23875 ($SD = 33932$) for unrelated targets. Again, all the pairs of conditions were matched: alternatives and related (t(48) = -0.048, p = 0.96), alternatives and unrelated (t(47) = -0.01, p = 0.99), related and unrelated (t(47) = 0.04, p = 0.97). Additionally, cosine similarity scores were calculated for each pair of noun in the sentence (X) and target, employing the LSAfun package with the dewak100k lsa Wordspace (Günther et al., 2015). Both alternatives ($M = .60$, $SD = .21$), and related targets ($M = .57$, $SD = .18$) were significantly more similar to the noun in the sentence than unrelated targets ($M = .13$, $SD = .10$; respectively $t(47) = 14.02$, $p < .001$ and $t(47) = 14.91$, $p < .001$). Alternatives and related targets did not differ significantly ($t(47) = 0.70$, $p = .49$).

As usual in priming studies, the mean difference between words and non-words was positive (17.15 ms), although this difference was not reliable (two-sided *t*-test: $t(53) = -1.504$, $p = .139$). Further analyses were run on the experimental items, i.e. those trials in which the target was an existing word. The data were analyzed with linear mixed effect models using the *lme4* package in *R* (Bates, 2005). A Model including the effect of Prime-Target Relation explained the data significantly better than the null model ($\chi^2(2) = 9.26, p < .05$). Both the alternative words and the semantically related words differed significantly from the unrelated words ($\beta = -35.02$, $t(2389) = -2.97$, $p < .05$ and $\beta = -24.77$, $t(2392) = -2.10$, $p < .05$ respectively). On the other hand, a model with Polarity as fixed effect did not provide any improvement over the null model ($\chi^2(1) = 0.40$, $p = .53$).

Figure 2.2

Pilot Study - Mean RTs by Relation and Polarity.



Note: Error bars represent ± SE of the means.

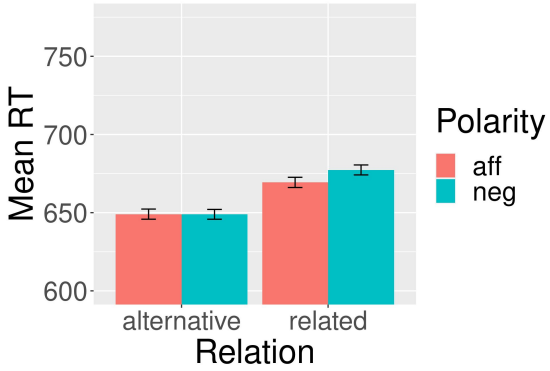
In order to test for the interaction of Relation and Polarity, unrelated words were omitted. A null model, which included the two fixed factors Relation and Polarity, as well as items and subjects as random factors, was compared to a model additionally including the interaction effect. The interaction did not improve the model significantly ($\chi^2(1) = 1.21$, $p = .27$). The mean RTs per condition are shown in Figure 2.2.

Although the interaction was not significant, the results are numerically in line with our hypothesis. We therefore ran a power analysis on the pilot data to determine a sample size that would reliably detect such an effect. A power analysis with 1000 simulations ($\alpha = .05$) resulted in 90.5% power when employing 240 subjects and 96 items.

In Experiment 1, data were collected until 240 usable participants (i.e. not excluded during the data cleaning phase) were reached. In order to select plausible alternatives we ran a cloze task prior to the main study. For this task we chose 100 concrete high frequency nouns (e.g. *melon*). One hundred subjects were instructed to complete sentences such as *Das ist keine Melone, das ist ...* (i.e. *This is not a melon, it's ...*) with either just a noun (e.g. *Ananas*) or an indefinite article plus a noun (e.g. *eine Ananas*). From the results we selected 96 nouns together with one of their frequent cloze answers to be used

Figure 2.3

Experiment 1 - Mean RTs by Relation and Polarity.



Note: Error bars represent ± SE of the means.

as plausible alternative in the main experiment.

As in the pilot, we controlled for length and frequency between targets across relations. Differently from the pilot, a match between alternatives and related targets could just not easily be achieved with these many items, as alternatives normally tend to score higher on similarity scores than semantically related non-alternatives. Although desirable, the match is not fundamental to the testing of our hypothesis.

The analysis was analogous to the analysis of the pilot. The mean difference between words and non-words was 31.35 ms, which was significant (two-sided t-test: $t(239) = -8.83$, $p < .001$). The model with Prime-Target Relation explained the data significantly better than the null model ($\chi^2(2) = 198.46, p < .001$). Again, both the alternatives and the semantically related words differed significantly from the unrelated words ($\beta = -39.93, t(20356) = -13.98$, $p < .001$ and $\beta = -15.40, t(20360) = -5.37$, $p < .001$ respectively). As in the pilot, a model with Polarity as fixed effect did not provide any improvement over the null model ($\chi^2(1) = 0.01$, $p = .91$). The interaction did not improve the comparison model significantly ($\chi^2(1) = 2.10$, $p = .15$). The mean RTs per condition are shown in Figure 2.3.

In our lab's study mentioned in the introduction, we found that completions to negative sentences (alternatives) were discernible from completions to affirmative sentences, in that they were significantly more similar to the negated entity. This was true across all sentential contexts employed, except for the *This* context (sentences of the form *This is [a/an]/[not a/an] X, it is [a/an]...*). To test whether the lack of an effect was due to the specifics of the *This* context, in Experiment 2 we changed the sentences of Experiment 1 to the form *There is [a/an]/[not a/an] X* (German *Dort ist [ein/e]/[kein/e] X*).

Additionally, to reduce noise, the response buttons for the lexical decision task were no longer counterbalanced across lists, ignoring the difference in reaction time between words and non-words, which was not fundamental to our hypothesis testing.

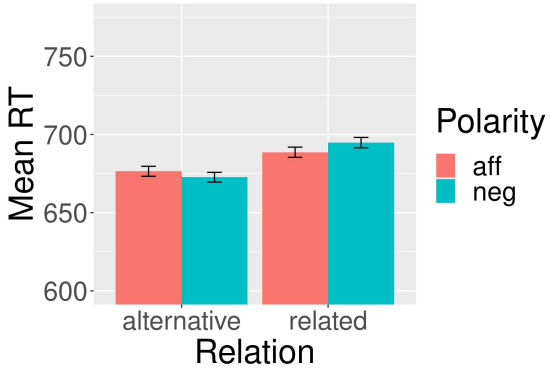
As in Experiment 1, data was collected until 240 usable participants were reached. We used the same material as in Experiment 1 but changed the sentences to the form *There is [a/an]/[not a/an] X* (German *Dort ist [ein/e]/[kein/e] X*). The procedure was also the same as in the Pilot and Experiment 1 except that the specific keys to press in this case were always the same across lists so the key for words was always *k* and for non-words always *d*.

For the analysis in this experiment, we used the same models as in the two prior experiments. Again, both the alternatives and the semantically related words differed significantly from the unrelated words ($\beta = -18.49$, $t(21254) = -10.06$, $p < .001$ and $\beta = -11.25$, $t(21257) = -3.96$, $p < .001$ respectively). As in the pilot and Experiment 1, a model with Polarity as fixed effect did not provide any improvement over the null model ($\chi^2(1) = 0.21$, $p = .65$). The interaction did not improve the model significantly either ($\chi^2(1) = 3.24$, $p = .07$). The mean RTs per condition are shown in Figure 2.4.

It struck us that nevertheless, numerically, the patterns of means consistently adhered to the hypothesis across experiments, especially considering that the targets presented in

Figure 2.4

Experiment 2 - Mean RTs by Relation and Polarity.



Note: Error bars represent ± SE of the means.

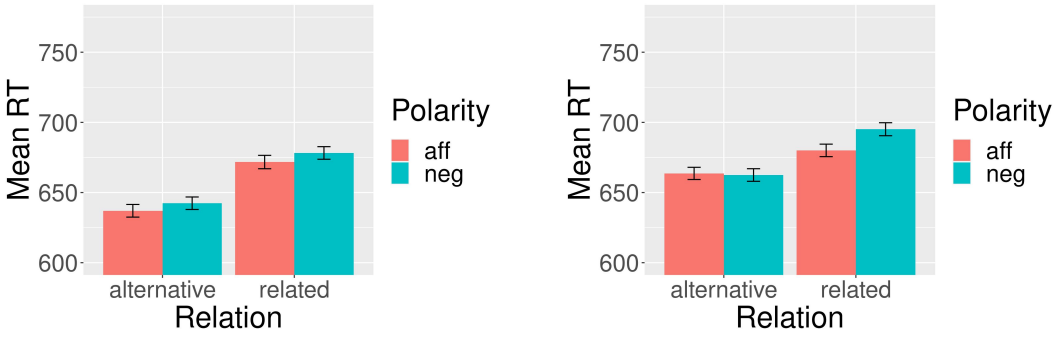
the affirmative and negative conditions are the same within each Relation condition. It is possible that the effect is present, but the pilot's sample overestimated it, whereas an even larger sample size would have been needed. Another consideration to make is that not all alternatives were *equally good* alternatives: some had higher cloze frequencies than others, and for 19 items cloze frequencies are not available as they were crafted by intuition. We can expect the alternatives with the highest cloze frequencies to be more likely candidates for an enhanced activation, as a larger proportion of subjects can be expected to prefer them. This possibility was explored in a post hoc analysis.

In order to explore the role of alternative *goodness*, we analysed the data of Experiment 1 and 2 whilst retaining only the items with a cloze frequency above the median (> 28). Experiment 1 showed no Relation x Polarity interaction ($\chi^2(1) = 0.02$, $p = .88$), but Experiment 2 did ($\chi^2(1) = 4.23$, $p < .05$): alternatives were facilitated after negation with respect to related, compared to the pattern of activation after affirmatives. The mean RTs per condition are shown in Figure 2.5. Therefore, power might not be the decisive issue, but rather the specifics of the sentential context and the goodness of alternatives.

Based on the results of the post hoc analysis of Experiment 2 on the *best* items, we

Figure 2.5

Post hoc Analysis - Mean RTs by Relation and Polarity.



Note: Mean RTs when retaining only the items with the highest cloze frequencies, in Experiment 1 (left panel) and in Experiment 2 (right panel). Error bars represent ± SE of the means.

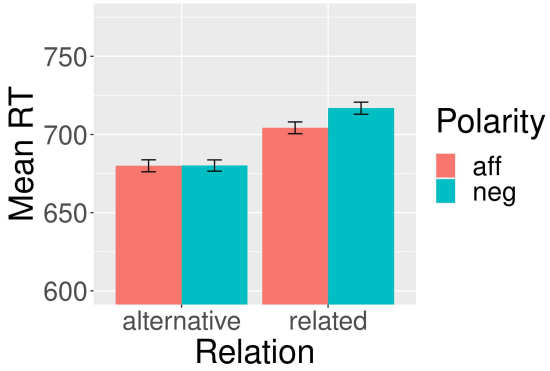
ran a third Experiment to investigate the expected interaction on a new sample. For that, we first conducted a power-analysis on Experiment 2 selecting only 48 best items in terms of the cloze frequencies obtained in our cloze task. The power analysis with 1000 simulations ($\alpha = .05$) resulted in 85.8% power when employing 540 subjects.

In contrast to the prior Experiments, the alternatives - but not the semantically related words - differed significantly from the unrelated words ($\beta = -32.45$, $p < .001$ and $\beta = -2.67$, $p < .41$, respectively). As in the prior Experiments, a model with Polarity as fixed effect did not provide any improvement over the null Model ($\chi^2(1) = 1.20$, $p = .27$). This time, the interaction reached significance level ($\chi^2(1) = 8.20$, $p < .05$). There was a significant Relation x Polarity interaction in the expected direction with alternatives being facilitated after negation with respect to related words, compared to the pattern of activation after affirmatives. The mean RTs per condition are shown in Figure 2.6.

Alternatives in the context of negation most commonly refer to the contrast between the negated and the expressed proposition (Repp & Spalek, 2021); accent has been placed on the time course of their access and integration in the mental model of the listener. Al-

Figure 2.6

Experiment 3 - Mean RTs by Relation and Polarity.



Note: Error bars represent ± SE of the means.

ternatives as the ones we refer to in the present study are traditionally investigated under the heading of *focus alternatives*, especially in relation to prosodically and syntactically marked focus. Different types of alternatives interact though, and disparate domains of alternatives have also been investigated jointly (Repp & Spalek, 2021). The current study suggests that negation functions as a (contrastive) focus marker, triggering focus alternatives without discourse context, and without prosodic marking. In fact, the pragmatic functions attributed to negation resonate with the notion of contrast delineated by Zimmermann et al. (2008): contrastive focus expresses the speaker's assumption that the listener does not expect the upcoming information; as such, it signals the need for a shift in the interlocutor's assumptions and an update of their common ground. Therefore, relevant alternatives are not simply dictated by semantic similarity, but by speakers' expectations on the status of their common ground, whereby semantic similarity is just a byproduct of the presentation of stimuli out of context. The discourse context is responsible for the restriction and therefore selection of the relevant alternatives. Orenes et al. (2014) showed that, after hearing a sentence such as *The figure is not green*, subjects ended up fixating the alternative (e.g. a blue figure) whenever only two concurrent alter-

natives were offered by a visual world (a green and a blue figure), but stayed fixated on the green figure when more alternatives were presented (e.g. green, blue, yellow, pink). They conclude that alternatives are activated when there are only two, but not when there are more than two. Whereas no generalized experience suggests that *blue* is a better alternative to *green* than *yellow* though, there is reason to assume that this is not always the case whenever more than one alternative is available (e.g. some entities can be widely agreed upon to be *better alternatives* than others). As a more general criterion, our study suggests that the prominence (rather than the number) of potential alternatives might be the decisive factor determining the activation. [2]

2.4 Using Circles Games to Investigate the Referential Use of Negation

As discussed, conversational negation displays pragmatic peculiarities and is not fully interchangeable with affirmation even when the amount of information conveyed is equated (e.g. Wason, 1961). Pragmatically felicitous contexts for negation have been identified, and include rejecting/correcting a false presupposition and referring to an exception. Nevertheless, studies on the spontaneous production of negation seem to show that it can follow general pragmatic principles such as informativity and relevance, without necessarily implying the rejection of false beliefs or reference to an exception (Beltrán et al., 2008; Nordmeyer & Frank, 2015; Watson, 1979). The focus of these studies has been specifically on the role of classical Gricean principles underlying successful communication (such as informativity and relevance), which are primarily concerned with information content and the derivation of conversational implicatures. In the present

[2] Reprinted from the previously published article Capuano et al. (2023) with permission from Springer Nature. See Appendix B for full manuscript.

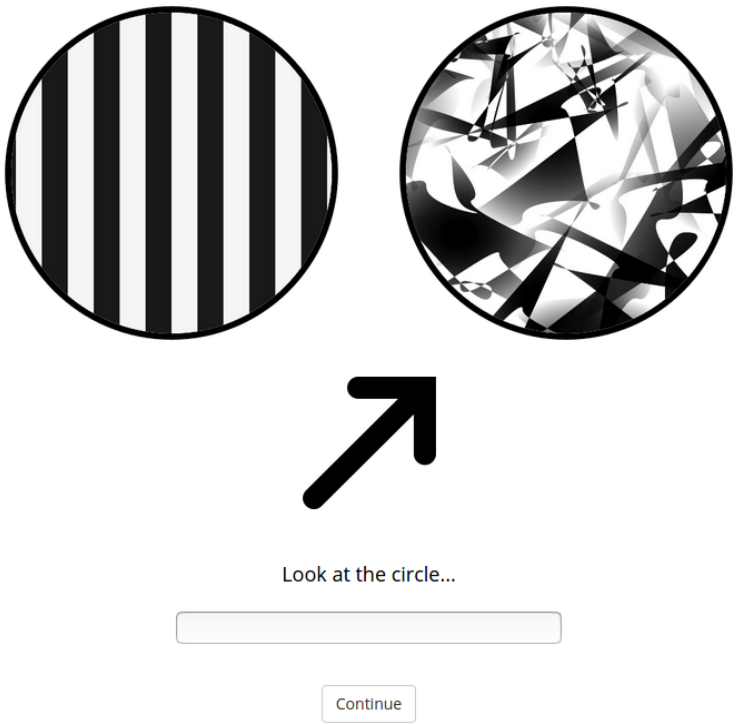
study we focused instead on the role of a more general principle of effort economy underlying a wider range of human behaviours (Zipf, 1949), understood - in the case of linguistic communication - as the preservation of the cognitive effort of the speaker to retrieve and produce an utterance (for a discussion of economy as a broader principle, see Carston, 2005). We examined whether negation production can be modulated by speaker economy, when the more economical form is not decidedly expected to be the affirmative option. In other words, can negation production be modulated alone by the ease of production with respect to an alternative affirmation that serves the same communicative purpose?

Subjects were presented with pairs of circles filled in varying patterns. For each pair, one circle was filled in a pattern that is easily nameable, the other with a pattern estimated hard to name. The circles were all displayed in the same size. An arrow would point at one of the two circles. Depending on the circle the arrow was pointed to, the item would vary on Target Difficulty (*easy* vs. *hard* to name). For an example stimulus, see Figure 2.7. Subjects were asked to complete the prompt *Look at the circle ...* so as to identify the circle being pointed at by the arrow: asserting a property of the target circle would be as informative as negating the property of the concurrent circle for the sake of identifying the correct one. The expectation is that, as long as it can fulfill the purpose of the speech act (identifying the correct circle), negation can be produced strategically when it more economically serves the purpose (e.g. *Look at the circle with no stripes* for Figure 2.7).

In the first three experiments economy was operationalised as the general ease to retrieve a description that would uniquely identify a referent. The items were 18 pairs of circles (Figure 2.7 is an example item for the first three experiments). The first three experiments (from now *Circle Games 1, 2* and *3*) differed in the delay of presentation of the arrow that identified the referent (0, 5000 and 500 ms respectively). It was hypothesised

Figure 2.7

Example Item of Circles Games.



Note: Example of an item in the *hard* to name Target Difficulty condition. On the left, the *easy* to name circle. On the right, the *hard* to name circle.

that longer delays would increase the availability of a description for the context, with respect to which a negative description of the referent could be formulated, therefore leading to a higher proportion of negative productions.

Negation was never produced when an affirmative description of the referent was *easy* to formulate, but it was produced strategically - even though less often than affirmation - when the affirmative description was *hard* and the same communicative purposes could be fulfilled with a more economical negative expression (Circles Game 1: $\chi^2(1) = 61.69$, $p < .001$, Circles Game 2: $\chi^2(1) = 71.34$, $p < .001$, Circles Game 3: $\chi^2(1) = 31.4, p < .001$). See Tables 2.6, 2.7, 2.8 for the results. The variations in the delay of presentation of the arrow did not produce any significant difference in the results. The hypothesis that

Table 2.6

Circles Game 1 - Number of Completions by Polarity and Difficulty.

	-	+	Total
easy	0	254	254
hard	48	164	212
Total	48	418	466

Table 2.7

Circles Game 2 - Number of Completions by Polarity and Difficulty.

	-	+	Total
easy	0	277	277
hard	54	173	227
Total	54	450	504

Table 2.8

Circles Game 3 - Number of Completions by Polarity and Difficulty.

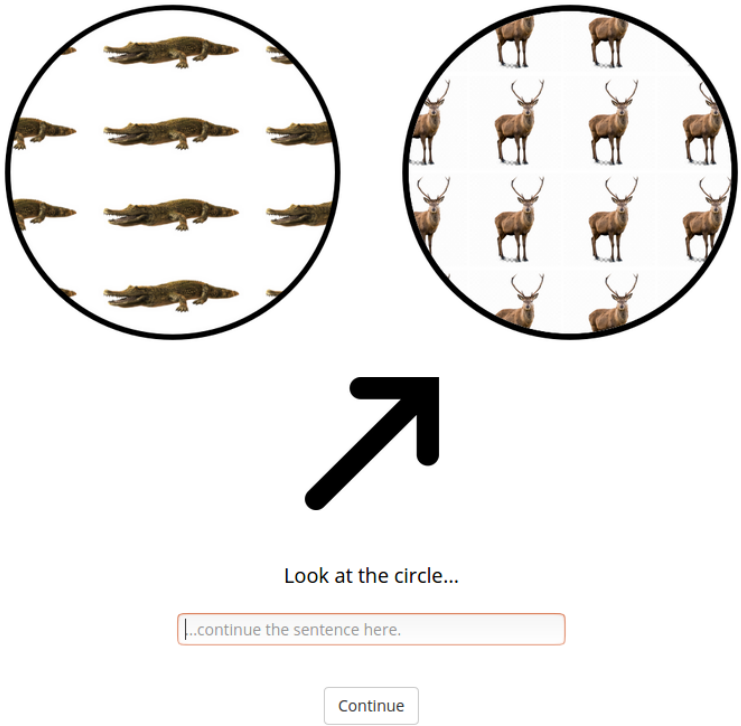
	-	+	Total
easy	0	186	186
hard	30	151	181
Total	30	337	367

longer delays would increase the proportion of negative productions was not confirmed and further investigation is required to identify the underlying cognitive processes and their time course.

In the fourth experiment economy was operationalised as the length of the description. The same 18 items from Circles Games 1-3 were employed and denoted as *pattern* items. Additionally, another 18 *length* items were created. These were also pairs of circles filled in a pattern, but the patterns were all repetitions of concrete entities. In the case of the *length* items, the naming difficulty of the two circles specifically differed in the length of the term needed to denote the entities in the circle: one was filled with a pattern of an entity with a long name, one with a pattern of an entity with a short name. See

Figure 2.8

Example of a Length Item.



Note: Example of a *length* item in the *easy* to name Target Difficulty condition. On the left, the *hard* to name circle (i.e. *crocodile*). On the right, the *easy* to name circle (i.e. *deer*).

Figure 2.8 for an example of *length* item. Referents were chosen such that they had high depictability, and such that a corresponding depiction could be found, that was associated to the referent as univocably as possible (e.g. we avoided words such as *skyscraper*, whose depictions could easily be identified simply by *building*).

The proportion of negative completions to the *pattern* items appears to be significantly higher than that to *length* items. The proportion of negative completions though does not seem to differ by length of description, and it lies almost at 0 (1 negative observation in the *easy* and 2 in the *hard* condition). See Table 2.9 for the results of Circles Game 4.

We additionally ran an exploratory analysis of indices of production effort such as sentence length and writing time across all experiments. The analysis reveals that: *easy*

Table 2.9

Circles Game 4 - Number of Completions by Item Type.

		-	+
length	easy	1	142
	hard	2	142
pattern	easy	0	130
	hard	14	115

Table 2.10

Exploratory Analysis - Completion Length and Writing Time in the Main Experiments.

	Circles Game 1	**Circles Game 2**	**Circles Game 3**	**Circles Game 4**
Completion Length	easy <hard ***	easy <hard ***	easy <hard ***	easy <hard ***
Completion Length	neg = aff	neg = aff	neg = aff	neg <aff *
Writing Time	easy <hard ***	easy <hard ***	easy <hard ***	easy <hard ***
Writing Time	neg = aff	neg <aff *	neg = aff	neg <aff *

circles always took less characters than *hard*; affirmative completions took as many as or more characters than negative; *easy* circles always took less time than *hard* circles to describe; affirmative completions took the same or more time than negative. See Table 2.10 for a concise summary. These exploratory results are compatible with the idea that, in those rare cases where subjects decided to employ negation, they did so because they had it available as a very advantageous option

We conclude that negation can and in fact is produced strategically on the basis of economy considerations. A clarification seems due relative to how the prevalence of affirmative sentences does not constitute an obstacle to this conclusion. The experiments modulated the economy of negative sentences relative to affirmative sentences: the effort to produce negatives is reduced in the hard condition, but this does not guarantee that negative sentences should be easier than affirmatives in that condition; it only guarantees that the difference in expected effort between negatives and affirmatives should be

greater in one condition than in the other. Although writing times seem to be in line with the idea that negatives were as economical as or more than affirmatives in the hard condition, that is not strictly speaking the main issue of interest. The study is rather concerned with the effect of the modulation of economy: if the effort to produce affirmative completions is increased, the probability to produce negatives should increase, as our results confirm. If, as it is, affirmation was still largely preferred, this might be due to economy being overcome by different pragmatic considerations: identifying a referent on the basis of a feature it possesses seems more pragmatically felicitous than identifying it on the basis of a feature it lacks. Although we assumed to have equated affirmatives and negatives on informativity, this might in fact not be the case. It has been argued that the informativity of a negative utterance is more accurately defined relative to a general discourse QUD, where *polar* QUDs render negation particularly informative (Xiang et al., 2020): in Nordmeyer and Frank (2014), the context highlighted the absence/presence of a specific property (i.e. having/not having apples) as particularly salient, implicitly suggesting an underlying polar QUD (*Does Bob have apples?*). In this sense, the large prevalence of affirmations produced in our study despite both negation and affirmation being apparently equated on informativity might be due to their deeper disparity in terms of underlying QUD (e.g. *How does the circle look like?* rather than *Does the circle have stripes?*) and therefore in terms of relevance. We can think of informativity as the update of prior beliefs: informative utterances reduce uncertainty in the state of affairs relative to a pertinent (i.e. relevant) question, with relevance being formalised by the concept of QUD. In our study, affirmatives and negatives were equated on informativity relative to the suggested communicative goal, in the sense that either would equally reduce the uncertainty in identifying the correct circle. Nevertheless, it seems like visual properties of the context might implicitly suggest other potential questions at issue.

In conclusion, the current study suggests that the effort of the speaker to retrieve and produce a negative utterance (i.e. strictly speaking speaker economy) contributes to its overall pragmatic felicity and modulates its production. Whereas the same is expected of any linguistic form, the production of negation is normally justified in terms of appropriateness of information content, seeing that, as a rule of thumb, negation is normally more effortful to produce than affirmation. Not only is speaker economy crucial, our results seem to suggest that greater pragmatic adequacy in terms of informativity is not a strictly necessary prerequisite for the production of negation: negation was spontaneously produced as it became a more and more economical option, even though affirmation was likely more informative for the question under discussion. [3]

[3] Reprinted from the previously published manuscript Capuano et al. (2022), © Copyright 2022, by permission of Informa UK Limited, trading as Taylor & Taylor & Francis Group, http://www.tandfonline.com. See Appendix C for full manuscript.

Chapter 3

General Discussion

The aim of this dissertation was to delve into the pragmatic peculiarities of negation. Conversational negation does not always behave like the logical operator. On the one hand, it is asymmetrical with respect to affirmation, the two not being interchangeable even in cases where they logically should; at the basis of this asymmetricality is the observation that they fulfil different pragmatic functions: negation is often associated with the rejection/correction of a previously held assumption. On the other hand, this specificity of negation results in the selection of only a few plausible alternatives: negating something does not suggest, like in the case of the logical operator, that any member of the complement set is an equally likely alternative. This specific aspect of conversational negation was the main focus of investigation of the thesis.

In Study 1, we examined whether the nature of the alternative scenarios that can be inferred from negation is consistent with the pragmatic peculiarities attributed to negation, and whether this is recognizably the case. In fact, if negation specifically advances similar alternatives, its behaviour should be discernible from affirmation. In a series of cloze tasks, we compared completions to negative sentences suggesting alternativehood

(e.g. *This is not ...., it is...* .) and structurally similar affirmative sentences (e.g. *This is ...., and that is ...* .). Completions to negative sentences appeared to be significantly more similar to the noun presented in the sentence than in the case of affirmative sentences. The only exception was represented by the *This is ....* . sentential context, which did non differ in completions with respect to polarity. Overall, these results were taken to support the idea that negation specifically suggests very similar alternatives, being in fact even more restrictive than a conjuncting affirmative context. The case of the *This* context might constitute an exemplary instance of affirmation conveying a contrast, a function which is probably not exclusive - only more specific - to negation.

Study 2 was designed to probe the activation of plausible alternatives during negation comprehension. Similarly to what is found in the case of other constructions for which the notion of alternative is relevant (e.g. sentences with focused items), negation would be expected to facilitate the activation of plausible alternatives, compared to the pattern of activation levels triggered by affirmation. A series of priming studies was run, where affirmative and negative sentences functioned as primes. The patterns of activation of nouns in different relations to the noun in the prime (alternatives vs. related) were compared between polarity conditions. All the experiments were numerically in line with the idea that negation facilitates the activation of alternatives. The facilitation though reached significance level only when very prominent alternatives were employed. The results suggest that negation behaves similarly to a contrastive focus marker, fostering an automatic search for plausible alternatives.

Finally, Study 3 looked at the preconditions for the production of negation referring to an alternative. Previous literature supports the claim that negation can be spontaneously produced even when no presupposition is being rejected. In fact, its production can be modulated by generic pragmatic principles such as relevance and informativity. Four

production experiments were conducted to test whether - relevance and informativity being equal - negation production additionally follows a principle of least effort. The results suggest that negation can in fact be produced based on economy considerations: negative sentences were generated significantly more often when a concurrent affirmative sentence was judged as harder to produce.

3.1 Integrating the Results in the Literature

The literature on negation has focused, over time, on different aspects that make negation unique and inform us on how it is processed. Negation has often proved to be an excellent test candidate given its structural peculiarities and its status as the most prototypical abstract, propositional procedure in language processing. The current results are relevant - although in unequal measure - to a number of points raised in the literature on negation, but also to general pragmatic models of language processing and the role of economy in language. In the following, we will try and integrate these outcomes in the literature by showing how they contribute to responding to some of the most relevant questions that have been raised on negation.

3.1.1 One- vs. Two-step Models of Negation

As seen in Chapter 1, it was proposed that negation does not have to be represented in a propositional format, but could be encoded as the sequence of representations of the non-factual (negated) state of affairs and the factual state of affairs. For example, processing a sentence like *The door is not closed* might involve the representation of a closed door, followed by an open door (*two-step model*). Alternatively, the *fusion model* suggests that when the alternative state of affairs is available (e.g. with binary

negation), comprehenders can 'jump' to the factual representation without the need to go through the representation of the negated state (Anderson et al., 2010; Giora, 2006; Orenes, 2021; Tian et al., 2010). Support to the two-step model is additionally given by the fact that, in order to decrease the processing costs of negation, it is not enough to use contrast markers like *contrary to expectations*; negation is only facilitated when the expectation to be denied can be derived in the prior context (see e.g. Albu et al., 2021), therefore when 'anticipating' the first step. Moreover, evidence that negation processing does not benefit from additional preparatory time in comparison to affirmation seems to be more in line with a two-step model of negation processing, postulating a first necessary representation of the negated state of affairs (Dudschig & Kaup, 2020a). What is mostly debated within this literature is whether the first step is a necessary condition for comprehension. Less attention has been given to the more or less mandatory nature of the second step. Nevertheless, it has been proposed - and evidence is in line with the idea that - the representation of the second step might depend on the kind of negation involved (non-binary vs. binary respectively). Whereas binary negation unambiguously introduces the factual state of affairs (*not closed* implies *open*), in the case of non-binary negation the listener might simply not have enough information to grant the simulation of the factual state of affairs (Du et al., 2014; Mayo et al., 2004; Orenes et al., 2014).

Study 2 puts forward the idea that, rather than a difference in terms of binary vs. non-binary negation, reaching the representation of the factual state of affairs might depend more generally on the plausibility (and therefore availability) of an alternative scenario. An alternative to a negated entity might not be picked deterministically but rather probabilistically. In fact, our sentences always involved non-binary negation, and differed only in the goodness of alternatives, approximated by cloze frequencies. Only when considering negations with very plausible alternatives did the results clearly show the activation

of an alternative ($\approx$ factual) state of affairs.

3.1.2 Activation Levels After Negation

The belief that negation could be represented explicitly as an operator over a proposition gave initially life to the idea that negation could function as an activation-reducing operator with respect to its scope. MacDonald and Just (1989) first looked at activation levels after negation. Their main finding was that negation inhibited the activation of the negated entity but not of its associates (at least not to a significant level).

Our results tell a different story. Study 2 suggests that the activation of the associates of a negated entity might depend on the more specific kind of semantic relationship the associate bears on the negated entity. We found semantic associates that are also good alternatives with respect to the negated entity (e.g. *wolf* with respect to *dog*) to be activated preferentially with respect to semantic associates that are not good alternatives (e.g. *bone* with respect to *dog*). The preferential activation was estimated from an interaction effect of sentence polarity and semantic relationship, but it is hard to discern whether the effect reflects an inhibition of related concepts or rather a facilitation of alternatives: the differences in reaction times between polarity conditions might be due to differences in sentence complexity. The question deserves further follow-up investigation and might require an inspection of the time course of these activation levels.

3.1.3 Conversational Negation and the Logical Operator

The divergent behaviour of linguistic particles like *and*, *if*, *not* and *or* from their logical counterparts was long debated in philosophy. Grice's theory of conversational implicatures (Grice, 1975) offered a new perspective on this longstanding dispute. Grice suggests that, beyond literal meaning, utterances can covey meaning which is generated

from the assumption that conversation is a cooperative act. Specifically, the speaker is assumed to be following four conversational maxims: Quantity (≈ 'be informative'), Quality (≈ 'be truthful'), Relation (≈ 'be relevant') and Manner (≈ 'be intelligible/-clear'). As an example, linguistic *or* tends to take on an exclusive reading because it is assumed that, if a speaker utters *A or B*, they do not believe or have enough knowledge that both A and B apply, otherwise they would have uttered *A and B*. Therefore, while the semantics of these particles might in fact align with the logical operators, their complex behaviour and meanings in natural language seem to be acquired by means of pragmatic principles.

As a main exponent of Neo-Gricean pragmatics, Horn applies similar ideas specifically to negation. Neo-Gricean accounts mostly diverge from the work of Grice by an attempt to refine or reduce the maxims. Closer to a linguistic perspective, Horn argues that the gricean maxims (excluding Quality, which is considered irreducible) are traceable to two basic principles: the Q principle (say as much as you can, given Quality and R) and the R principle (say no more than you must, given Q). The two principles reflect the tensions towards hearer's and speaker's economy respectively, which interact and restrain each other. Horn defines a division of pragmatic labour between the two forces: unmarked forms (shorter, more lexicalised, more economical in general) tend to be associated with a standard meaning through the R principle, whereas marked forms (more verbose, complex and less lexicalised) usually Q-implicate non-standard meanings. As anticipated in 1.3, negation is often Q-restricted to convey the correction of an expectation, else an affirmation would be employed fulfilling different informativity requirements.

Looking at Study 1, our results confirm that negation is in fact employed very differently from the corresponding logical operator: if it were not, any element of the comple-

ment set could have been selected as an alternative to a negated entity. Instead, speakers expressed a clear preference for very similar entities. Moreover, the way speakers used negation reflects the sense predicted by the Gricean accounts: the marked form, negation, tends to take on a marked meaning, with a behaviour discernible from that of affirmation.

3.1.4 Negation is Harder to Process than Affirmation

Across the literature, negation consistently produced longer latencies and processing times than affirmation, which do not seem exclusively attributable to differences in sentence length. What was later observed though is that, when negation is presented in a pragmatically felicitous context, these differences in processing times are largely reduced. The question arose from the consideration that, whereas experiments often presented negative sentences out of context, negation in everyday usage is often employed in contexts where it corrects a false presupposition.

Our results are in line with the idea that this is in fact a particularly felicitous context for negation: in our production tasks from Study 1, speakers completed negative sentences in accordance with this reading. Indeed, the fact that they selected alternatives that are very similar to the negated entity renders the context of the false presupposition more plausible, because it increases the likelihood that the two entities could have been erroneously mixed up.

3.1.5 Correction of a Presupposition as Generalized Implicature

Neo-Gricean Levinson employs two similar principles to Horn's (Q and I for Informativeness), and adds a third principle, Manner:

Indicate abnormal, non-stereotypical situations by using marked expressions that contrast with those you would use to describe the corresponding normal, stereotypical situ-

ations.

Fundamentally, Levinson observes the existence of preferred interpretations and inferences that are relatively invariant across contexts. It is suggested that powerful heuristics give rise to these preferred interpretations: an order of applicability of the principles is introduced to avoid clashes ($Q > M > R(I)$). In this sense, *generalized conversational implicatures* are preferred, stereotyped implicatures.

The understanding of negation as a correction of a previously held assumption seems to be a preferred interpretation. This is supported by the fact that, even in relatively unconstrained production tasks like those of Study 1, which employed minimalistic sentential contexts and did not in any way suggest that completions should hint at a corrective reading, actual completions suggest that this was nevertheless the default function attributed to negation.

3.1.6 Automaticity

The current results additionally link to the literature on negation and automaticity. Researchers have asked whether negation is processed automatically, in the sense that it might be fully processed unintentionally even in cases where the task does not require it. Results in this sense are mixed.

On the one hand, Deutsch et al. (2006) have investigated the issue through valence and found no evidence for a quick, automatic integration of negation. They show that prime nouns with a positive or negative valence facilitate the processing of a target with the same valence, independently of whether the prime was negated or not (e.g. *a/no war* both primed *disease*).

On the other hand, Deutsch et al. (2009) changed the targets from words to Chinese ideograms, which were to be judged on visual pleasantness. An interaction of polarity

and noun valence suggests that speakers did in fact integrate negation fast and unintentionally. Similarly, Armstrong and Dienes (2013) observed priming effects on subliminally processed negated and non-negated words, consistent with an automatic integration of the negation operator. Nevertheless, the effect in Deutsch et al. (2009) was not found when the subjects were distracted by a digit memory task, suggesting that negation integration is fast and automatic only when enough cognitive resources are available, which might answer to why at the end of full length sentences subjects are often not yet done processing negation.

Study 2 seems to reinforce the view that negation can be integrated automatically: even though our task did not involve coming up with a plausible factual state of affairs, participants automatically engaged in a search for a plausible alternative. Moreover, this was the case in the context of short sentences, suggesting that automatic integration of negation can in principle take place beyond a simple noun phrase.

3.1.7 Negation and General Cognitive Mechanisms

More recently, the literature has focused on the possibility that negation might employ more general cognitive mechanisms well-known in other areas of cognitive science. Negation might share mechanisms with action control, specifically those involved in motor inhibition (Beltrán et al., 2021; Beltrán et al., 2019; de Vega et al., 2016; Liu et al., 2020). Alongside, Dudschig and Kaup (2018) show how negation processing displays similarities to non-linguistic conflict detection and adaptation (see also Dudschig & Kaup, 2020b; Dudschig, Kaup, Svaldi, et al., 2021).

Similarly, in a way, we looked at whether a principle of least effort, governing different aspects of human behaviour apart from language more specifically, plays a role in negation production. Whereas the tension between speaker's economy and listener's economy

might have molded the generalised preference for negation to correct a presupposition, more contingent pragmatic considerations might also have an effect on modulating online negation production. Stress was put mostly on other pragmatic factors governing the spontaneous production of negation, such as relevance or informativity. We investigated more directly pure linguistic economy in terms of sentence complexity/availability and length. Study 3 demonstrates that simple linguistic economy can, in fact, modulate negation production, even overriding the general preference to use affirmation to identify a referent (i.e. it is generally preferred to describe a referent by positively referring to its properties, rather than to the lack of a property with respect to a concurrent referent).

3.2 Conclusions

In conclusion, although the current dissertation focused on the thematic of negation and alternatives, the results add up to the discussion on a variety of the main issues raised in the negation literature. Negation seems to specifically foster a default corrective reading, coming to the point of automatically engaging a search for plausible alternatives. Notwithstanding this generalized reading, its production can be modulated by more online pragmatic considerations, such as pure differences in economy.

Overall, the results demonstrate a behaviour of negation that is in line with Gricean accounts of language use, confirming the explanatory force and wide applicability of a pragmatic treatment of negation. Taken together, the current studies speak for the relevance of the underlying pragmatic theories and engage a step forward in the understanding of the specificities of negation in natural language.

Bibliography

Albu, E., Tsaregorodtseva, O., & Kaup, B. (2021). Contrary to expectations: Does context influence the processing cost associated with negation?. *Journal of Psycholinguistic Research*, *50*(6), 1215–1242. https://doi.org/10.1007/s10936-021-09799-8

Anderson, S. E., Huette, S., Matlock, T., & Spivey, M. (2010). On the temporal dynamics of negated perceptual simulations. In F. Parrill, V. Tobin, & M. B. Turner (Eds.), *Meaning, form, & body* (pp. 1–20). CSLI Stanford, CA.

Armstrong, A.-M., & Dienes, Z. (2013). Subliminal understanding of negation: Unconscious control by subliminal processing of word pairs. *Consciousness and Cognition*, *22*(3), 1022–1040. https://doi.org/10.1016/j.concog.2013.06.010

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! a systematic comparison of context-counting vs. context-predicting semantic vectors. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 238–247. https://doi.org/10.3115/v1/p14-1023

Bates, D. (2005). Fitting linear mixed models in r. *R news*, *5*(1), 27–30.

Beltrán, D., Liu, B., & de Vega, M. (2021). Inhibitory mechanisms in the processing of negations: A neural reuse hypothesis. *Journal of Psycholinguistic Research*, *50*(6), 1243–1260. https://doi.org/10.1007/s10936-021-09796-x

Beltrán, D., Morera, Y., García-Marco, E., & Vega, M. d. (2019). Brain inhibitory mechanisms are involved in the processing of sentential negation, regardless of its content. evidence from eeg theta and beta rhythms. *Frontiers in Psychology*, *10*, Article 1782. https://doi.org/10.3389/fpsyg.2019.01782

Beltrán, D., Orenes, I., & Santamaría, C. (2008). Context effects on the spontaneous production of negation. *Intercultural Pragmatics*, *5*(4), 409–419. https://doi.org/10.1515/iprg.2008.020

Braun, B., & Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes*, *25*(7-9), 1024–1043. https://doi.org/10.1080/01690960903036836

Bruni, E., Tran, N.-K., & Baroni, M. (2014). Multimodal distributional semantics. *Journal of Artificial Intelligence Research*, *49*, 1–47. https://doi.org/10.1613/jair.4135

Capuano, F., Dudschig, C., Günther, F., & Kaup, B. (2021). Semantic similarity of alternatives fostered by conversational negation. *Cognitive Science*, *45*(7), Article e13015. https://doi.org/10.1111/cogs.13015

Capuano, F., Dudschig, C., & Kaup, B. (2022). Using circles games to investigate the referential use of negation. *Language, Cognition and Neuroscience*, *38*(1), 65–76. https://doi.org/10.1080/23273798.2022.2083645

Capuano, F., Sorg, T., & Kaup, B. (2023). Activation levels of plausible alternatives in conversational negation. *Memory & Cognition*, 1–12. https://doi.org/10.3758/s13421-023-01434-2

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, *82*(1), 45–73. https://doi.org/10.1037/h0076248

Carston, R. (2005). Relevance theory, grice and the neo-griceans: A response to laurence horn's 'current issues in neo-gricean pragmatics'. *Intercultural Pragmatics*, *2*(3), 303–319. https://doi.org/doi:10.1515/iprg.2005.2.3.303

Chase, W. G., & Clark, H. H. (1972). Mental operations in the comparison of sentences and pictures. In F. Parrill, V. Tobin, & M. B. Turner (Eds.), *Cognition in learning and memory*. John Wiley & Sons.

Chomsky, N. (1957). *Syntactic structures*. De Gruyter Mouton. https://doi.org/10.1515/9783112316009

Clark, H. H. (1976). *Semantics and comprehension*. De Gruyter Mouton. https://doi.org/10.1515/9783110871029

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*(3), 472–517. https://doi.org/10.1016/0010-0285(72)90019-9

Clark, H. H., & Chase, W. G. (1974). Perceptual coding strategies in the formation and verification of descriptions. *Memory & Cognition*, *2*(1), 101–111. https://doi.org/10.3758/bf03197499

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. Harcourt Brace Jovanovich Ltd.

Cornish, E. R. (1971). Pragmatic aspects of negation in sentence evaluation and completion tasks. *British Journal of Psychology*, *62*(4), 505–511. https://doi.org/10.1111/j.2044-8295.1971.tb02065.x

Dahl, Ö. (1979). Typology of sentence negation. *Linguistics*, *17*(1-2), 79–106. https://doi.org/10.1515/ling.1979.17.1-2.79

De Swart, H. (2009). *Expression and interpretation of negation: An ot typology* (Vol. 77). Springer Dordrecht. https://doi.org/10.1007/978-90-481-3162-4

De Villiers, J. G., & Flusberg, H. B. T. (1975). Some facts one simply cannot deny. *Journal of Child Language*, *2*(2), 279–286. https://doi.org/10.1017/s0305000900001100

Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: Negation as reflective operation. *Journal of Personality and Social Psychology*, *91*(3), 385–405. https://doi.org/10.1037/0022-3514.91.3.385

Deutsch, R., Kordts-Freudinger, R., Gawronski, B., & Strack, F. (2009). Fast and fragile: A new look at the automaticity of negation processing. *Experimental Psychology*, *56*(6), 434–446. https://doi.org/10.1027/1618-3169.56.6.434

de Vega, M., Morera, Y., León, I., Beltrán, D., Casado, P., & Martín-Loeches, M. (2016). Sentential negation might share neurophysiological mechanisms with action inhibition. evidence from frontal theta rhythm. *Journal of Neuroscience*, *36*(22), 6002–6010. https://doi.org/10.1523/JNEUROSCI.3736-15.2016

Du, P., Liu, D., Zhang, L., Hitchman, G., & Lin, C. (2014). The processing of contradictory and non-contradictory negative sentences. *Journal of Cognitive Psychology*, *26*(4), 461–472. https://doi.org/10.1080/20445911.2014.903957

Dudschig, C., & Kaup, B. (2018). How does "not left" become "right"? electrophysiological evidence for a dynamic conflict-bound negation processing account. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(5), 716–728. https://doi.org/10.1037/xhp0000481

Dudschig, C., & Kaup, B. (2020a). Can we prepare to negate? negation as a reversal operator. *Journal of Cognition*, *3*(1), 32. https://doi.org/10.5334/joc.119

Dudschig, C., & Kaup, B. (2020b). Negation as conflict: Conflict adaptation following negating vertical spatial words. *Brain and Language*, *210*, Article 104842. https://doi.org/10.1016/j.bandl.2020.104842

Dudschig, C., & Kaup, B. (2021). Pictorial vs. linguistic negation: Investigating negation in imperatives across different symbol domains. *Acta Psychologica*, *214*, Article 103266. https://doi.org/10.1016/j.actpsy.2021.103266

Dudschig, C., Kaup, B., Liu, M., & Schwab, J. (2021). The processing of negation and polarity: An overview. *Journal of Psycholinguistic Research*, *50*(6), 1199–1213. https://doi.org/10.1007/s10936-021-09817-9

Dudschig, C., Kaup, B., Svaldi, J., & Gulewitsch, M. D. (2021). Negation processing in children with adhd: The generic problem of using negation in instructions. *Journal of Psycholinguistic Research*, *50*(6), 1309–1320. https://doi.org/10.1007/s10936-021-09789-w

Dudschig, C., Mackenzie, I. G., Maienborn, C., Kaup, B., & Leuthold, H. (2019). Negation and the n400: Investigating temporal aspects of negation integration using semantic and world-knowledge violations. *Language, Cognition and Neuroscience*, *34*(3), 309–319. https://doi.org/10.1080/23273798.2018.1535127

Eifermann, R. R. (1961). Negation: A linguistic variable. *Acta Psychologica*, *18*, 269–278. https://doi.org/10.1016/0001-6918(61)90016-6

Engelkamp, J., & Hörmann, H. (1974). The effect of non-verbal information on the recall of negation. *Quarterly Journal of Experimental Psychology*, *26*(1), 98–105. https://doi.org/10.1080/14640747408400391

Fillenbaum, S. (1966). Memory for gist: Some relevant variables. *Language and Speech*, *9*(4), 217–227. https://doi.org/10.1177/002383096600900403

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry Jr, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*(4), 400–409. https://doi.org/10.1111/j.1469-8986.1983.tb00920.x

Giora, R. (2006). Anything negatives can do affirmatives can do just as well, except for some metaphors. *Journal of Pragmatics*, *38*(7), 981–1014. https://doi.org/10.1016/j.pragma.2005.12.006

Givón, T. (1978). Negation in language: Pragmatics, function, ontology. In *Pragmatics* (pp. 69–112). Brill. https://doi.org/10.1163/9789004368873_005

Glenberg, A. M., Robertson, D. A., Jansen, J. L., & Johnson-Glenberg, M. C. (1999). Not propositions. *Cognitive Systems Research*, *1*(1), 19–33. https://doi.org/10.1016/s1389-0417(99)00004-2

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill. https://doi.org/10.1163/9789004368811_003

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun - An R package for computations based on latent semantic analysis. *Behavior Research Methods*, *47*, 930–944. https://doi.org/10.3758/s13428-014-0529-0

Günther, F., Dudschig, C., & Kaup, B. (2016). Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies. *Quarterly Journal of Experimental Psychology*, *69*(4), 626–653. https://doi.org/10.1080/17470218.2015.1038280

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033. https://doi.org/10.1177/1745691619861372

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2-3), 146–162. https://doi.org/10.1080/00437956.1954.11659520

Hörmann, H. (1971). Semantic factors in negation. *Psychologische Forschung*, *35*(1), 1–16. https://doi.org/10.1007/BF00424470

Horn, L. (1984). Toward a new taxonomy for pragmatic inference: Q-based and r-based implicature. In D. Schiffrin (Ed.), *Meaning, form, and use in context: Linguistic applications* (pp. 11–42). Georgetown University Press.

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

Husband, E. M., & Ferreira, F. (2016). The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience*, *31*(2), 217–235. https://doi.org/10.1080/23273798.2015.1083113

Jones, M. N., Kintsch, W., & Mewhort, D. J. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534–552. https://doi.org/10.1016/j.jml.2006.07.003

Jones, S. (1968). Instructions, self-instructions and performance. *Quarterly Journal of Experimental Psychology*, *20*(1), 74–78. https://doi.org/10.1080/14640746808400130

Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 244–253. https://doi.org/10.1016/s0022-5371(71)80051-8

Kaup, B., Ludtke, J., & Zwaan, R. A. (2005). Effects of negation, truth value, and delay on picture recognition after reading affirmative and negative sentences. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *27*. https://escholarship.org/uc/item/19s068vb

Kaup, B., & Dudschig, C. (2020). Understanding negation: Issues in the processing of negation. In V. Déprez & M. T. Espinal (Eds.), *The oxford handbook of negation* (pp. 635–655). Oxford University Press.

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics*, *38*(7), 1033–1050. https://doi.org/10.1016/j.pragma.2005.09.012

Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Lüdtke, J. (2007). Experiential simulations of negated text information. *Quarterly Journal of Experimental Psychology*, *60*(7), 976–990. https://doi.org/10.1080/17470210600823512

Kaup, B., & Zwaan, R. A. (2003). Effects of negation and situational presence on the accessibility of text information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(3), 439–446. https://doi.org/10.1037/0278-7393.29.3.439

Kruszewski, G., Paperno, D., Bernardi, R., & Baroni, M. (2016). There is no logical negation here, but there are alternatives: Modeling conversational negation with distributional semantics. *Computational Linguistics*, *42*(4), 637–660. https://doi.org/10.1162/coli_a_00262

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211–240. https://doi.org/10.1037/0033-295x.104.2.211

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, *20*(1), 1–31.

Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics*, *4*, 151–171. https://doi.org/10.1146/annurev-linguistics-030514-125254

Liu, B., Wang, H., Beltrán, D., Gu, B., Liang, T., Wang, X., & de Vega, M. (2020). The generalizability of inhibition-related processes in the comprehension of linguistic negation. erp evidence from the mandarin language. *Language, Cognition and Neuroscience*, *35*(7), 885–895. https://doi.org/10.1080/23273798.2019.1662460

Lüdtke, J., & Kaup, B. (2006). Context effects when reading negative and affirmative sentences. *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, *27*, 1735–1740.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208. https://doi.org/10.3758/bf03204766

MacDonald, M. C., & Just, M. A. (1989). Changes in activation levels with negation. *Journal of experimental psychology: Learning, memory, and cognition*, *15*(4), 633–642. https://doi.org/10.1037/0278-7393.15.4.633

Mauri, C., & Van der Auwera, J. (2012). Connectives. In K. Allan & K. M. Jaszczolt (Eds.), *The cambridge handbook of negation* (pp. 377–401). Cambridge University Press. https://doi.org/10.1017/cbo9781139022453.021

Mayo, R., Schul, Y., & Burnstein, E. (2004). "i am not guilty" vs "i am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology*, *40*(4), 433–449. https://doi.org/10.1016/j.jesp.2003.07.008

Miestamo, M. (2007). Negation - an overview of typological research. *Language and Linguistics Compass*, *1*(5), 552–570. https://doi.org/10.1111/j.1749-818x.2007.00026.x

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407. https://doi.org/10.3758/bf03195588

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. https://doi.org/10.1111/j.1467-9280.2008.02226.x

Nordmeyer, A., & Frank, M. C. (2015). *Negation is only hard to process when it is pragmatically infelicitous* [Unpublished manuscript].

Nordmeyer, A., & Frank, M. (2014). A pragmatic account of the processing of negative sentences. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *36*. https://escholarship.org/uc/item/95j3b18w

Orenes, I. (2021). "looking at" negation: Faster processing for symbolic rather than iconic representations. *Journal of Psycholinguistic Research*, *50*(6), 1417–1436. https://doi.org/10.1007/s10936-021-09797-w

Orenes, I., Beltrán, D., & Santamaria, C. (2014). How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and Language*, *74*, 36–45. https://doi.org/10.1016/j.jml.2014.04.001

Repp, S., & Spalek, K. (2021). The role of alternatives in language. *Frontiers in Communication*, *6*. https://doi.org/10.3389/fcomm.2021.682009

Roberts, C. (1996/2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and pragmatics*, *5*(6), 1–69. https://doi.org/10.3765/sp.5.6

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics*, *1*(1), 75–116. https://doi.org/10.1007/bf02342617

Rück, F., Dudschig, C., Mackenzie, I. G., Vogt, A., Leuthold, H., & Kaup, B. (2021). The role of predictability during negation processing in truth-value judgment tasks. *Journal of Psycholinguistic Research*, *50*(6), 1437–1459. https://doi.org/10.1007/s10936-021-09804-0

Silvennoinen, O. O. (2020). Contrastive negation: Constructional variation within and across languages. *Neuphilologische mitteilungen*, *120*(2), 207–212.

Tian, Y., Breheny, R., & Ferguson, H. J. (2010). Why we simulate negated information: A dynamic pragmatic account. *Quarterly Journal of Experimental Psychology*, *63*(12), 2305–2312. https://doi.org/10.1080/17470218.2010.525712

Tian, Y., Ferguson, H., & Breheny, R. (2016). Processing negation without context–why and when we represent the positive argument. *Language, Cognition and Neuroscience*, *31*(5), 683–698. https://doi.org/10.1080/23273798.2016.1140214

Tottie, G. (1982). Where do negative sentences come from? *Studia Linguistica*, *36*(1), 88–105. https://doi.org/10.1111/j.1467-9582.1982.tb00714.x

Wales, R., & Grieve, R. (1969). What is so difficult about negation? *Perception & Psychophysics*, *6*(6), 327–332. https://doi.org/10.3758/bf03212785

Wason, P. C. (1961). Response to affirmative and negative binary statements. *British Journal of Psychology*, *52*(2), 133–142. https://doi.org/10.1111/j.2044-8295.1961.tb00775.x

Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, *4*(1), 7–11. https://doi.org/10.1016/s0022-5371(65)80060-3

Wason, P. C. (1972). In real life negatives are false. *Logique et Analyse*, *15*(57/58), 17–38.

Wason, P. C., & Jones, S. (1963). Negatives: Denotation and connotation. *British Journal of Psychology*, *54*(4), 299–307. https://doi.org/10.1111/j.2044-8295.1963.tb00885.x

Watson, J. M. (1979). Referential description by children in negative form. *British Journal of Psychology*, *70*(2), 199–204. https://doi.org/10.1111/j.2044-8295.1979.tb01676.x

Xiang, M., Kramer, A., & Nordmeyer, A. (2020). An informativity-based account of negation complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(10), 1857–1867. https://doi.org/10.1037/xlm0000851

Yan, M., & Calhoun, S. (2019). Priming effects of focus in mandarin chinese. *Frontiers in Psychology*, *10*, 1985. https://doi.org/10.3389/fpsyg.2019.01985

Zimmermann, M., et al. (2008). Contrastive focus and emphasis. *Acta Linguistica Hungarica*, *55*(3-4), 347–360. https://doi.org/10.1556/aling.55.2008.3-4.9

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.

Appendix A

Study 1: Capuano et al. (2021)

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 45 (2021) e13015

ISSN: 1551-6709 online
DOI: 10.1111/cogs.13015

Semantic Similarity of Alternatives Fostered by Conversational Negation

Francesca Capuano, Carolin Dudschig, Fritz Günther, Barbara Kaup

Department of Psychology, University of Tübingen

Received 18 January 2021; received in revised form 1 June 2021; accepted 5 June 2021

Abstract

Conversational negation often behaves differently from negation as a logical operator: when rejecting a state of affairs, it does not present all members of the complement set as equally plausible alternatives, but it rather suggests some of them as more plausible than others (e.g., "This is not a dog, it is a wolf/*screwdriver"). Entities that are semantically similar to a negated entity tend to be judged as better alternatives (Kruszewski et al., 2016). In fact, Kruszewski et al. (2016) show that the cosine similarity scores between the distributional semantics representations of a negated noun and its potential alternatives are highly correlated with the negated noun-alternatives human plausibility ratings. In a series of cloze tasks, we show that negation likewise restricts the production of plausible alternatives to similar entities. Furthermore, completions to negative sentences appear to be even more restricted than completions to an affirmative conjunctive context, hinting at a peculiarity of negation.

Keywords: Negation; Pragmatics; Language production; Distributional semantics; Alternatives; Cloze tasks

1. Introduction

Negation is a universal property of natural language. Despite the variations, every human language has devices to achieve truth value reversal (Horn, 1989; Miestamo, 2007). While negation seems to be pervasive, it is often associated with processing and integration difficulties (Deutsch, Gawronski, & Strack, 2006; Dudschig, Mackenzie, Maienborn, Kaup, &

Correspondence should be sent to Francesca Capuano, Department of Psychology, University of Tübingen, Schleichstr. 4, 72076 Tübingen, Germany. E-mail: francesca.capuano@uni-tuebingen.de

Leuthold, 2019; Fischler, Bloom, Childers, Roucos, & Perry, 1983). Early psycholinguistic studies on negation focused on the differences between affirmative and negative sentences, systematically detecting an increase in processing times for the negative sentences (for an overview, see Kaup & Dudschig, 2020). Wason (1961), for example, presented subjects with sentences such as *Twenty-four [is]/[is not] an [even]/[odd] number* to be verified for truth value. Negation appeared harder to process than affirmation well beyond what can be explained by variations in sentence length. From a logical perspective, the two sentences should be equivalent as they convey membership to one of two mutually exclusive classes; therefore, they could in principle be used interchangeably. The pitfall is the assumption of a purely logical treatment of negation whereas in fact, in natural language, negation is often not used in its strictly logical sense. After all, the usage of negation where a logically equivalent affirmation is applicable must be justified as carrying additional inferences (Grice, 1975; Moeschler, 1992). As noted early on by Wason (1965), the studies overlooked this issue: by presenting the sentences out of context, they ignored the specific pragmatic usages of negation. For example, a sentence like *24 is not odd*, presented in isolation, might be harder to process than the affirmative *24 is even*, simply because it does not fulfill the pragmatic functions normally attributed to negation.

Indeed, subsequent studies confirmed that the processing difficulties of negation are at least partly attributable to its specific pragmatic nuances: if presented in a pragmatically felicitous context, the processing costs of negation are significantly reduced (e.g., Glenberg, Robertson, Jansen, & Johnson-Glenberg, 1999; Johnson-Laird & Tridgell, 1972; Lüdtke & Kaup, 2006; Nieuwland & Kuperberg, 2008; Wason, 1965). Pragmatically felicitous contexts include, among others, instances in which negation is used to refer to an exception (e.g., Cornish, 1971; Nordmeyer & Frank, 2015; Wason, 1965) or to something that is easily confusable with a contrast set (De Villiers & Flusberg, 1975). Wason (1965) presented subjects with stimuli in which one of eight otherwise identical items differed from the rest in color (e.g., eight circles, seven in blue and one in red). Subjects were asked to complete affirmative and negative sentences referring to the color of one of the eight items (e.g., *Circle n.4 [is]/[is not]____*). The difference in response times between negative and affirmative sentences referring to the exceptional item was significantly smaller than the difference between negative and affirmative sentences referring to one of the similar items (*exceptionality hypothesis*). De Villiers and Flusberg (1975) similarly expected the difference in completion times between negative and affirmative sentences to be smaller when the exceptional referent belonged to the same more general category as the rule members (e.g., *A cow is not a horse* vs. *A shoe is not a flower*). They showed children to be increasingly sensitive to confusability with age. These findings are all in line with the general idea that negation is often used to indicate deviations from expectations and/or to correct a false presupposition (Clark & Clark, 1977; Givón, 1978; Horn, 1989; Wason, 1965): comprehenders, in turn, seem to be sensitive to this pragmatic aspect, processing negation faster if it is used in line with this pragmatic principle.

A negative utterance therefore does not just convey the falsity of a statement, but additionally the belief of the speaker that the listener might presume the statement to be true. In this sense, linguistic negation does not just reject aspects of the described state of affairs, but might additionally give rise to the inference that an alternative scenario applies, justifying

the choice of its usage. For example, when hearing an utterance such as *This is not a dog*, in order to make sense of it as congruent with the violation of an expectation and/or with the correction of a presupposition, we might expect the referent of the utterance to be more likely a wolf rather than, for example, a screwdriver (Kruszewski et al., 2016). As Kruszewski et al. (2016) noted, it follows that logic-based approaches are not well suited to model this pragmatic aspect of conversational negation: rather than denoting the whole complement set indiscriminately, negation acts as a graded similarity function, defining a probability distribution over a restricted set of alternative scenarios.

Alternativehood implies some level of exchangeability that is contextually defined, therefore plausible alternatives are context-dependent. But although any member of the complement set of an entity could potentially be a plausible alternative given enough contextual pressure, some entities tend to be better alternatives than others, in that they are relevant across a large number of contexts. In this sense, the notion of alternativehood is tied to that of contextual similarity: as good alternatives are interchangeable with the negated entity, they can be expected to occur in similar contexts (Kruszewski et al., 2016). Starting from these considerations, Kruszewski et al. (2016) show that the plausibility of alternatives introduced by negation can be effectively modeled in terms of contextual similarity with respect to the negated entity through distributional semantic (DS) representations. Distributional semantic models (DSMs) are computational models deriving vectorial distributed representations of meaning out of information extracted from corpora (Landauer & Dumais, 1997; Lenci, 2018; Lund & Burgess, 1996). The intuition behind the models is that words that occur in similar contexts will tend to have similar meanings (Harris, 1954). For example, both *sea* and *ocean* will often be used in the same sentence as *water*, *fish*, or *sail*. For each word, the models return a high-dimensional vector that encodes information on which contexts the word occurs in. Words that occur in similar contexts will result in similar vectors. A cosine similarity score can be calculated between pairs of vectors to reflect the degree of similarity in meaning between the two corresponding words. The models effectively construct a semantic space where the distance between words reflects their semantic distance. Even though with limitations and criticisms, DSMs are able to reflect cognitive similarity measures such as those derived from priming studies or similarity judgments (Bruni, Tran, & Baroni, 2014; Günther, Dudschig, & Kaup, 2016; Günther, Rinaldi, & Marelli, 2019; Jones et al., 2006; Lenci, 2008).

Kruszewski et al. (2016) show cosine similarity scores to be a good predictor of plausibility of alternatives fostered by negation. They collected human plausibility ratings of sentences in the two following forms:

(1) a. This is not X, it is Y.
b. There is no X here, but there is Y.

X were 50 randomly sampled nouns. Y were nouns collected among co-hyponyms, hyponyms, hypernyms, nearest neighbors, functionally similar items, visually similar nouns, free associates, and unrelated nouns with respect to X. Sentences in which the noun (Y) was more similar to the negated entity (X) were rated by subjects as more plausible. The cosine similarity scores between the DS representations of the negated nouns and the alternatives were highly correlated with the human ratings.

Starting from this evidence that a negated sentence is rated as more acceptable the higher the presented alternative scores in DSM measures of similarity, the current study was designed to answer two key questions that remained open: (a) is this phenomenon also seen in production, in that speakers tend to produce alternatives to negations that score high in DSM similarity measures? (b) In how far is this higher similarity specific to negation, if we compare it to standard language use?

Regarding the first point, if negation elicits very similar alternatives, then these are the words participants should be expected to produce when prompted to do so. Additionally, production tasks partly overcome the limitations of selecting item material (as opposed to comprehension tasks): although the sentential contexts are given (e.g., *There is no X here, but there is*____), subjects are free to produce any alternative Y they deem appropriate without limitations on the solutions space, possibly resulting in a clearer overview of actual language usage. In fact, a comprehension task entails that the author selects the alternatives. Such a selection of the item material might alter the results by assuming a reality different from the true population. Therefore, switching to a production task might potentially uncover a selection bias (e.g., Forster, 2000). As for the second point, the previous study did not compare negated and affirmative sentences. It is therefore not informative for the assumption that negation specifically comes with the presentation of very similar alternatives: the possibility persists that affirmative sentences elicit even higher similarities. In fact, the correlation between the similarity of nouns within a sentence and plausibility ratings might hold true in general (e.g., people might also prefer an affirmative sentence such as *This is a dog and that is a wolf* over *This is a dog and that is a screwdriver*): without a baseline comparison condition, we will not be able to tell whether negation specifically prompts very similar alternatives beyond what can be suggested by, for example, lexical association and contiguity. It is therefore necessary to investigate the question of whether specifically negation results in the production of similar scenarios, by comparing the results to an affirmative conjunctive context.

We constructed minimal negative sentences suggesting alternativehood and had participants complete them in a series of cloze tasks. As a comparison baseline, we used affirmative sentences with the *and* connector. The sentences were kept structurally as similar as possible to the negative sentences. We expected the baseline to already give rise to the production of very similar entities, as *and* is typically associated with coding co-occurrence of state of affairs (Mauri & Van der Auwera, 2012), and therefore can be expected to suggest contiguity. Nevertheless, as alternativehood goes beyond contiguity, additionally implying the substitutability of entities, we expected even higher similarity scores for entities produced in negative sentences. Showing this would suggest that negation indeed has the specific function to prompt very similar alternatives.

2. General method

As negated nouns (X) we always used the 50 nouns employed by Kruszewski et al. (2016) (e.g., *goat*; see Supporting Information Appendix A for the complete list of the 50 nouns). In order to get a more complete overview of the distribution of semantic similarity in natural language, we additionally employed four reference points to assess whether the

words produced by participants can actually be considered "similar" on an absolute level, as would be expected if negation indeed involves the presentation of very similar alternatives. The reference points were same for all experiments and were calculated with respect to these 50 nouns. All the cosine similarity scores were derived from the same vector space employed in Kruszewski et al. (2016), with the help of the LSAfun package (Günther et al., 2015). The vector space is the one that performed closest to empirical results throughout a series of evaluation tasks (e.g., semantic relatedness, synonym detection) in Baroni et al. (2014).

The first reference point is the mean similarity score of the 50 nouns with their respective closest neighboring noun in the semantic space. The second reference point was calculated as the mean of the average cosine similarity score of each noun with their free associates, normalized by the number of participants producing the free associate response. Free associates and their production frequencies were extracted from the University of South Florida (USF)-free association norms (Nelson et al., 2004), which covered 42 of the 50 nouns. The third point is the mean of the average similarity score of each noun with all the nouns it co-occurs with within the constructions [Noun] [*and*/*or*/,] [optional: *the*/*a*] [Noun], weighted by the co-occurrence frequencies. The fourth reference point is the mean of the average similarity score of each noun with all the nouns it co-occurs with within a sentence, weighted by the co-occurrence frequencies. Co-occurrences for the third and fourth reference points were extracted from the same corpus that was used to construct the semantic space. The last point is the mean of the average cosine similarity score of each noun with 100 randomly sampled nouns.

Responses to the cloze tasks were crowdsourced via Amazon Mechanical Turk. The tasks were estimated to take each 10–15 min and were rewarded with $1.50. The location of the workers was set in the United States. The survey link instructions requested the workers to be native English speakers. At the beginning of the task, participants gave informed consent and were further inquired on their native language. Workers who had already participated in one of the tasks were excluded from further participation. The experiments were programmed with the aid of jsPsych (De Leeuw, 2015) and participants ran them on their chosen device using standard Internet browsers. Participants typed their responses using their standard keyboard.

3. Experiment 1

3.1. Participants

Data were collected from 72 participants in Task 1a and another 72 in Task 1b. After data cleaning (details described in Section 3.4), 56 subjects remained in the final dataset in Task 1a (34 males and 22 females; age (in years): mean = 36.43, $SD = 9.4$) and 51 subjects in Task 1b (32 males, 18 females and 1 other; age: mean = 36.73, $SD = 11.89$).

3.2. Materials

The 50 nouns were presented in negative or affirmative sentences for completion in cloze tasks. For the negative sentences, we used the same *This* and *There* sentence types

as Kruszewski et al. (2016), deprived of the alternative. In Task 1a, participants were then requested to complete negative sentences such as

(2) a. There is no goat here, but there is____.
b. This is not a goat, it is____.

We then slightly adapted the negative sentences to turn them into affirmative sentences, while keeping them as superficially similar to the negative ones as we could. In Task 1b, participants were then requested to complete affirmative sentences such as

(3) a. There is a goat here, and there is____ there.
b. This is a goat, and that is____.

3.3. Procedure

Participants were presented with the following instructions:

In this study you will be presented with incomplete sentences. For each sentence, your task is to imagine a scenario in which you think the sentence is likely to be uttered, and complete it in a natural way.

The tasks were run in one block. Each participant was presented with all 50 nouns from Kruszewski et al. (2016), each randomly assigned either to the *This* or to the *There Sentence Type* condition . The order of presentation of the sentences was randomized.

3.4. Results

Participants who stated their native language is other than English were excluded from the analysis ($n = 0$). The completion strings were lowercased, stripped, and deprived of punctuation (except for "-"). Stopwords were removed. At this point, for simplicity, only answers consisting of one noun were considered (96.33% of the data points). All answers identical to the negated word were excluded. A cosine similarity score was calculated for each negated noun-completion noun pair. The score was derived from the best performing vector space from Baroni et al. (2014) using the LSAfun package (Günther et al., 2015). Completion nouns that were not found in the semantic space were not considered (6 out of 709 unique completions were not found in Task 1a, 1 out of 733 in Task 1b). All participants left with less than 15 answers per *Sentence Type* condition were excluded ($n = 37$). The cleaning procedure was identical for all experiments reported in this manuscript.

To test our hypothesis that negative sentences should receive completions with a higher similarity score with respect to the negated noun (as opposed to affirmative sentences), we employed linear mixed effect models using the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). The models will be described via R syntax (R Core Team, 2017). For this analysis, data from Task 1a and Task 1b were pooled, resulting in a dataset where *Polarity* (i.e., negative vs. affirmative sentences) is a between-subject factor. The following linear mixed effect model was applied:

$$Cosine \sim Polarity + (1|Item) + (1|Subject) \quad (1)$$

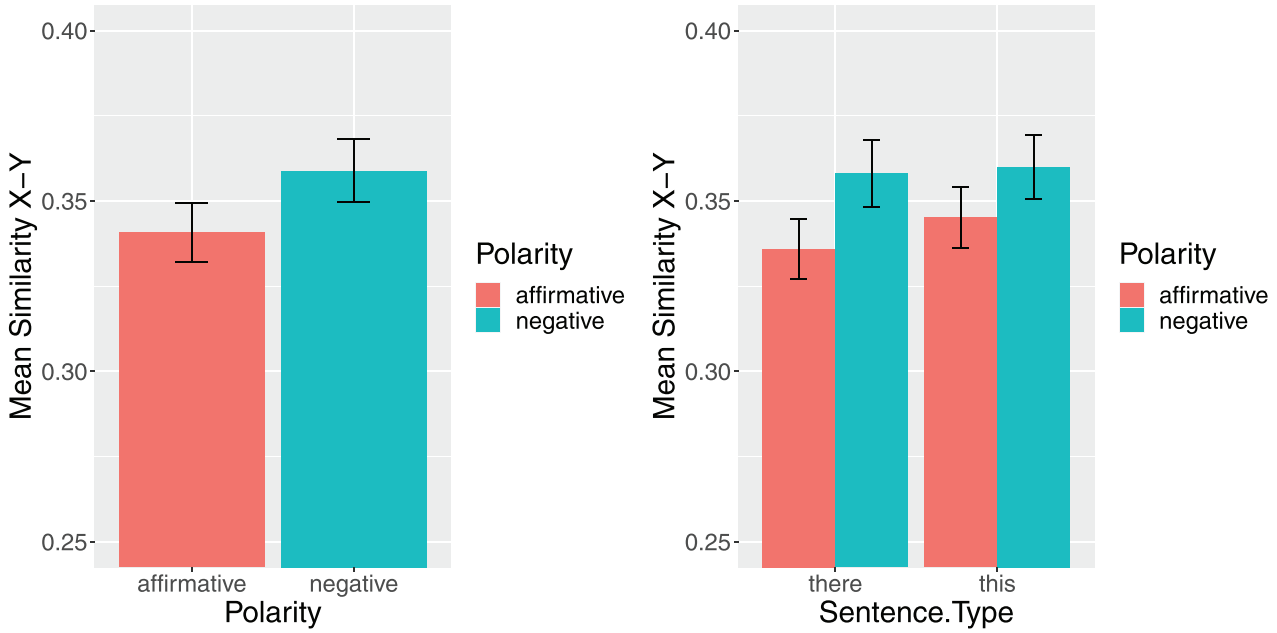


Fig. 1. Experiment 1—Mean cosine similarity scores (aggregated by subject) of affirmative and negative conditions, overall (left panel) and for each *Sentence Type* (right panel). Error bars represent $\pm SE$ of the means.

and compared to the baseline model

$$Cosine \sim 1 + (1|Item) + (1|Subject) \tag{2}$$

through a likelihood ratio test.

Model 1 does not explain the data significantly better than Model 2 ($\chi^2(1) = 2.08$, $p = .15$, $d = 0.13$), therefore *Polarity* does not seem to be a relevant factor in modeling the data. Cohen's d is calculated following Westfall, Kenny, and Judd (2014) (see Brysbaert & Stevens, 2018).

A comparison between Model 3

$$Cosine \sim Polarity \times \ Sentence \cdot Type + (1|Item) + (1|Subject) \tag{3}$$

and Model 4

$$Cosine \sim Polarity + Sentence \cdot Type + (1|Item) + (1|Subject) \tag{4}$$

results in a nonsignificant interaction of *Polarity* and *Sentence Type* ($\chi^2(1) = 2.23$, $p = .14$). A comparison between Model 5

$$Cosine \sim Sentence \cdot Type + (1|Item) + (1|Subject) \tag{5}$$

and Model 2 shows a nonsignificant effect of *Sentence Type* ($\chi^2(1) = 3.16$, $p = .08$); see Fig. 1 for the results.

Nevertheless, a trend in the expected direction can be observed in the data. As far as our knowledge goes, this study is the first of its kind and might be underpowered. In the following experiments, we manipulated *Polarity* within participants to increase the probability of detecting a potential effect.

4. Experiment 2

4.1. Participants

Data were collected from 98 participants. The cleaning procedure left us with 84 subjects (45 males, 37 females, and 2 others; age: mean = 38.15, $SD = 10.81$).

4.2. Materials

From Experiment 2 onward, the experimental design is within subjects: each participant is presented with all 50 nouns, each randomly assigned to either the affirmative or the negative condition. Each experiment corresponds to one task that employs a single sentence type.

In this task (Task 2), participants completed negative and affirmative *There*-type sentences, such as

(4) a. There is no goat here, but there is____.
b. There is a goat here, and there is____ there.

4.3. Procedure

In order to decrease the answers' rejection rate, participants were now explicitly instructed to complete the sentences with only one noun. The instructions were changed to

In this study you will be presented with incomplete sentences, such as "There is no camel here, but there is____." or "There is a camel here, and there is____ there." Your task is to complete each sentence in a natural, grammatical way. Please complete each sentence with only one noun (e.g., "water"), or a determiner + a noun (e.g., "a dune").

The task was run in one block. Each participant was presented with all the 50 items from Kruszewski et al. (2016), each randomly assigned either to the affirmative or the negative context. The order of presentation of the sentences was randomized.

4.4. Results

The data were analyzed by comparing the following linear mixed effect models:

$$Cosine \sim Polarity + (1|Item) + (1 + Polarity|Subject) \tag{6}$$

$$Cosine \sim 1 + (1|Item) + (1 + Polarity|Subject) \tag{7}$$

through a likelihood ratio test. The same two models will be referenced back throughout the paper as they were employed for the analysis of all the tasks from now on.

As expected in our initial hypothesis, *Polarity* plays a significant role in explaining the data ($\chi^2(1) = 13.17$, $p < .001$, $d = 0.11$), in that negative sentences receive more similar completions with respect to the noun at issue than affirmative sentences (see Fig. 2).

When looking at the results of Experiment 1 by *Sentence Type* (see Fig. 1), one might expect the effect to be stronger in the *There* context. In Experiment 3 we will test whether, employing a within-subject design, the effect emerges in the *This* context as well.

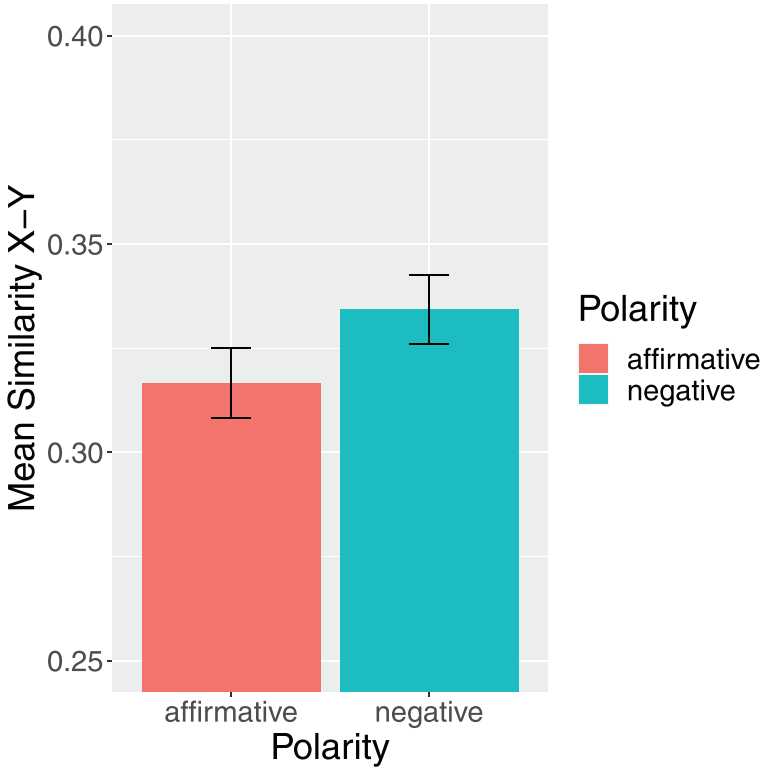


Fig. 2. *Experiment* 2—Mean cosine similarity scores (aggregated by subject) of affirmative and negative conditions. Error bars represent ±*SE* of the means.

5. Experiment 3

5.1. Participants

Data were collected until the same number of usable participants as in Experiment 2 was reached ($n = 84$; 53 males and 31 females; age: mean = 36.3, $SD = 9.42$). In total, data were collected from 104 participants.

5.2. Materials

In this task (Task 3), participants completed negative and affirmative *This* sentences, such as

(5) a. This is not a goat, it is____.
b. This is a goat, and that is____.

5.3. Procedure

The instructions were analogous to those in Experiment 2. The rest of the procedure was same as in Experiment 2.

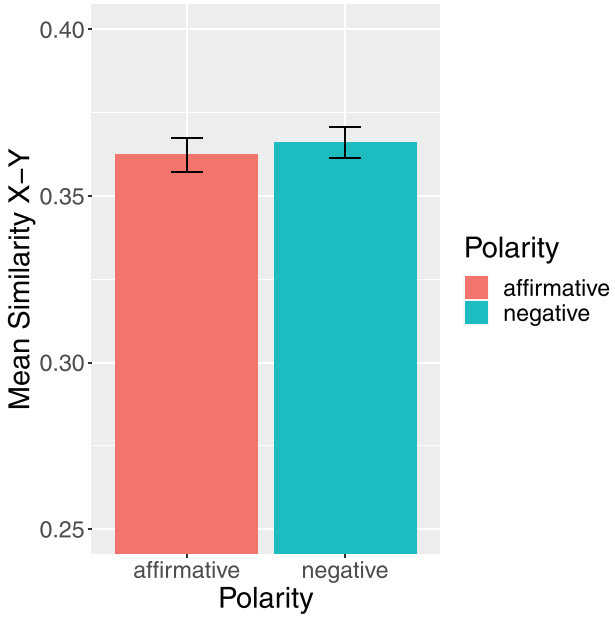


Fig. 3. *Experiment* 3—Mean cosine similarity scores (aggregated by subject) of affirmative and negative conditions. Error bars represent ±*SE* of the means.

5.4. Results

We fit again Model 6 and compared it to Model 7. This time, Model 6 did not explain the data better than Model 7 ($\chi^2(1) = 1.03$, $p = .31$, $d = 0.03$) (see Fig. 3). This could be interpreted as a ceiling effect: in the *This* contexts, completions produced to the affirmative sentences are already so similar to the noun at issue, that even more similar ones cannot be produced, simply because they are not accessible/do not exist. Alternatively, the presence of the adverbs *here* and *there* in the negative sentences of Experiment 2 might have driven the effect in the first place, rather than the presence of the negator, by evoking two separate scenarios for the two entities. In Experiment 4, we rendered the items of Experiment 3 analogous to the items in Experiment 2, by adding the adverb *here* and the conjunction *but* to the negative sentences, and the adverbs *here* and *there* to the affirmative sentences. We ran a power analysis on Experiment 2 to rule out a power issue, given that the same effect would be expected for the *This* contexts. In order to be sensible, power analysis needs an effect to be present. Therefore, we cannot base a power analysis on Experiment 3 (where we have no effect). We reached 0.89 power after 1000 simulations with 100 participants (Brysbaert & Stevens, 2018).

6. Experiment 4

6.1. Participants

Data were collected until 100 usable participants were reached (68 males, 31 females, and 1 other; age: mean = 39.15, *SD* = 11.81). In total, data were collected from 135 participants.

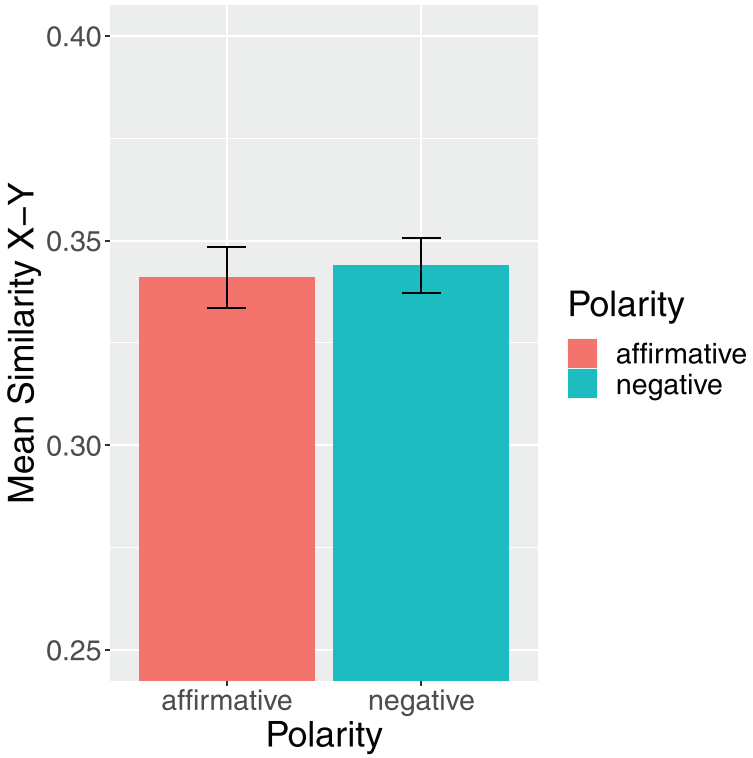


Fig. 4. *Experiment* 4—Mean cosine similarity scores (aggregated by subject) of affirmative and negative conditions. Error bars represent ±*SE* of the means.

6.2. Materials

Participants completed a slight variation of the negative and affirmative *This* sentences from Experiment 3. Analogously to the *There* sentences from Experiment 2, the adverb *here* and the conjunction *but* were added to the affirmative sentences, and both adverbs *here* and *there* were added to the negative sentences, such as

(6) a. This is not a goat *here*, *but* it is____.
b. This is a goat *here*, and that is____ *there*.

6.3. Procedure

Same as in Experiment 3, with the example sentences changed accordingly.

6.4. Results

Once again, Model 6 could not account for the data better than Model 7 ($\chi^2(1) = 1.86$, $p = .17$, $d = 0.03$) (see Fig. 4). This time the absence of the effect might be driven by the concessive meaning that *but* can assume in the negative sentences (e.g., Blakemore, 1989; Lang, 2008; Malchukov, 2004). This would render hypernyms, rather than co-hyponyms, more plausible completions (e.g., *This is not a goat, but it is an animal.*≈*This is not a goat, nevertheless it is an animal.*), therefore increasing the distance in similarity between the two

entities. In fact, cosine similarity scores are shown to favor symmetric relations, such as co-hyponymy over hypernymy (Lenci & Benotto, 2012).

7. Experiment 5

7.1. Participants

Data were collected until 100 usable participants were reached (59 males, 40 females, and 1 other; age: mean = 36.24, $SD = 10.63$). In total, data were collected from 139 participants.

7.2. Materials

We used the same sentences as in Experiment 4, but got rid of the *but* in the negative sentences:

(7) a. This is not a goat here, it is____.
b. This is a goat here, and that is____ there.

7.3. Procedure

Same as in Experiment 3.

7.4. Results

Again, a model including *Polarity* as a fixed effect does not explain the data any better than a model without *Polarity* ($\chi^2(1) = 0.48$, $p = .49$, $d = 0.02$) (see Fig. 5). We hypothesized that all the *This* affirmative contexts, similar to the negative contexts, specifically suggest a contrast between the two entities, therefore resulting in a null effect. The question will receive further consideration in Section 11. The question remains, whether the findings of Experiment 2 are generalizable to any other sentential context. We tested this in Experiments 6 and 7. New sentential contexts were chosen such that—similar to the *There* and the *This* contexts—they would not be too restrictive, allowing for a large answer space, which is not the case for many other potential constructions with too many selectional restrictions (e.g., a sentence like *I eat no* ____ limits the answer space to the food domain).

8. Experiment 6

8.1. Participants

Data were collected until 100 usable participants were reached (61 males and 39 females; age: mean = 35.67, $SD = 9.79$). In total, data were collected from 146 participants.

8.2. Materials

We now asked participants to complete different sentences, of the form

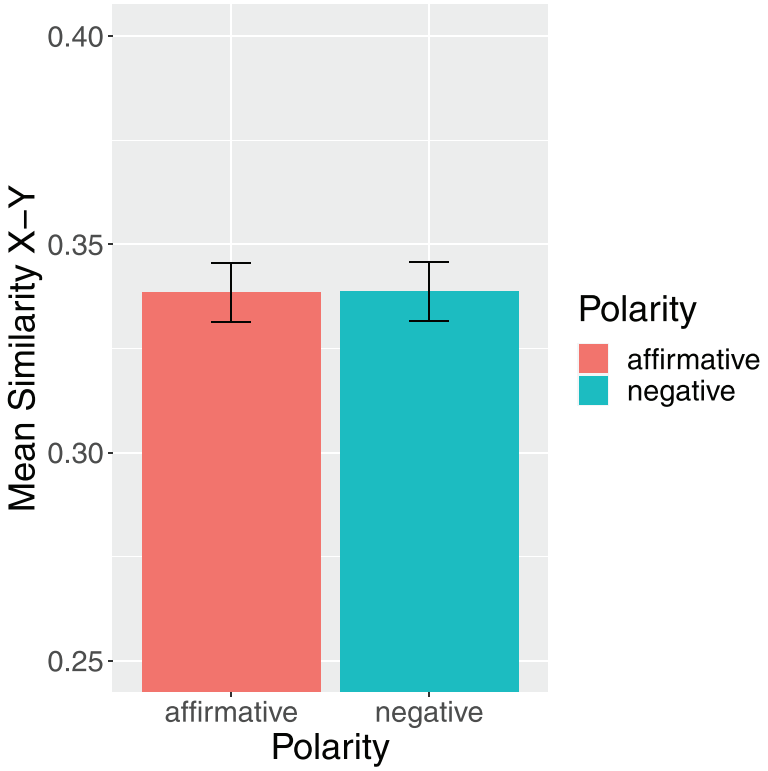


Fig. 5. *Experiment* 5—Mean cosine similarity scores (aggregated by subject) of affirmative and negative conditions. Error bars represent ±*SE* of the means.

(8) a. (Pronoun) see(s) no X, but (Pronoun) see(s)____.
b. (Pronoun) see(s) a X and (Pronoun) see(s)____.

The sentences were again kept superficially as similar to each other as we could. We employed all subject personal pronouns (*I*, *you*, *he*, *she*, *it*, *we*, *they*). The same pronoun was always used in both the first and second positions. An example item would be

(9) a. We see no goat, but we see____.
b. We see a goat and we see____.

8.3. Procedure

Same as in Experiment 3, with the addition that pronouns were also randomly sampled.

8.4. Results

Model 6 explains the data significantly better than Model 7 ($\chi^2(1) = 22.75$, $p < .001$, $d = 0.16$) (see Fig. 6). Experiment 7 was run to ensure that the effect in Experiment 6 indeed generalizes to other sentence types.

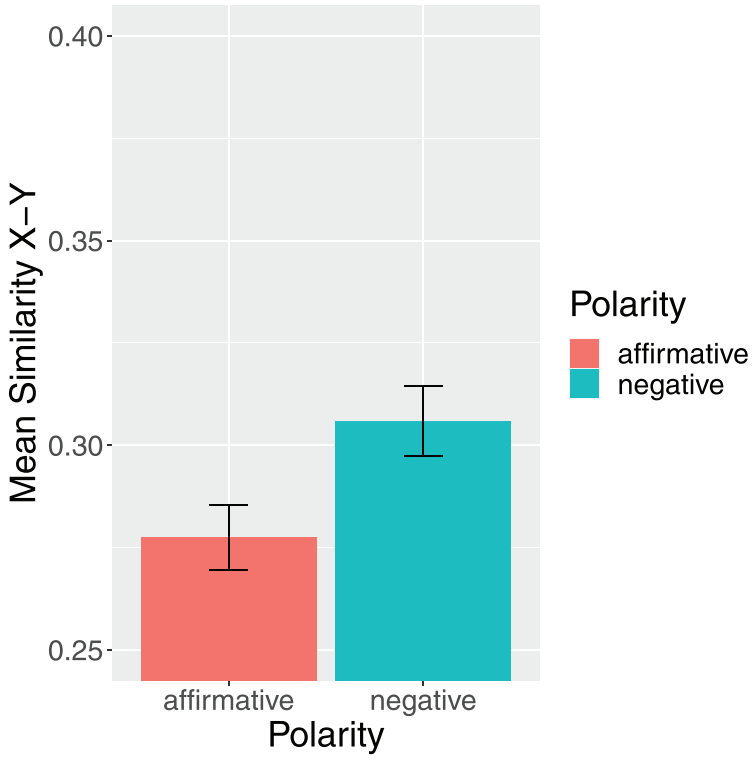


Fig. 6. *Experiment* 6—Mean cosine similarity scores (aggregated by subject) of affirmative and negative conditions. Error bars represent ±*SE* of the means.

9. Experiment 7

9.1. Participants

Data were collected until 100 usable participants were reached (49 males and 50 females; age: mean = 36.49, *SD* = 10.43). In total, data were collected from 164 participants.

9.2. Materials

Participants were asked to complete sentences of the form

(10) a. (Pronoun) want(s) no X, but (Pronoun) want(s)____.
b. (Pronoun) want(s) a X and (Pronoun) want(s)____.

The pronoun selection was akin to the one in Experiment 6. An example item would be

(11) a. We want no goat, but we want____.
b. We want a goat and we want____.

9.3. Procedure

Same as in Experiment 6.

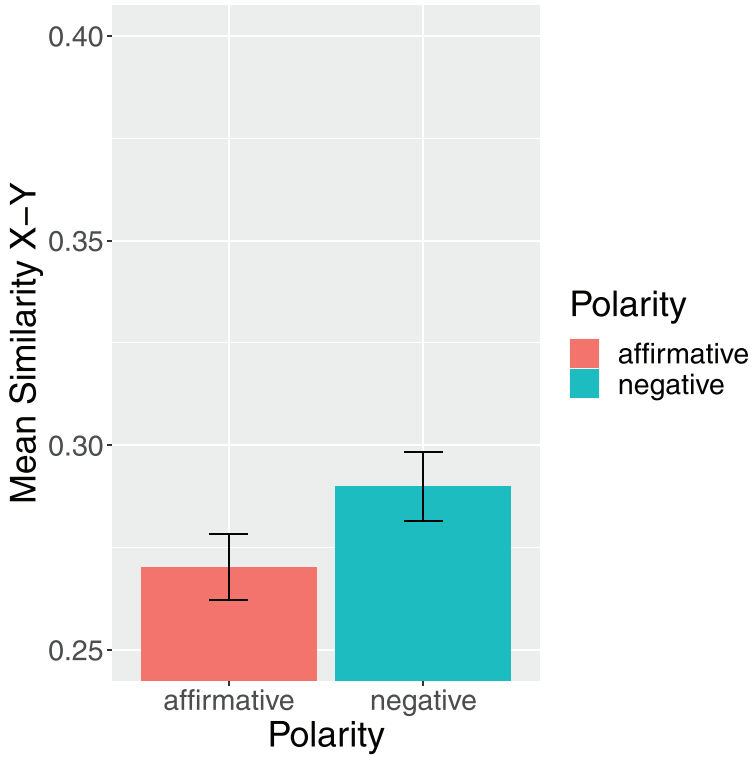


Fig. 7. *Experiment* 7—Mean cosine similarity scores (aggregated by subject) of affirmative and negative conditions. Error bars represent ±*SE* of the means.

9.4. Results

Model 6 explains the data significantly better than Model 7 ($\chi^2(1) = 29.77$, $p < .001$, $d = 0.15$) (see Fig. 7).

Raw data from all experiments, manuscript and analysis code can be found at https://osf.io/p762c/?view_only=8e6c73dd6f9d42848641b28f12863449. A summary of the results of all the experiments, together with the reference values, can be found in Fig. 8.

10. Post hoc analysis

For most of the sentence types employed ("there," "see," and "want" vs. "this"), the results show that negation restricts completions more than conjunction in an affirmative context. The restriction was defined in terms of semantic similarity. We additionally explored whether completions to affirmative and negative sentences differed on a relational level. Intuitively, one could expect a preference for completions in paradigmatic relationship (De Saussure, 2011) in the case of negation. Even in contexts that allow for the alternative to be in a syntagmatic relationship with the negated entity (e.g., *There is no dog, but there is a bone* vs. **This is not a dog, it is a bone*) the absence of an entity might still favor its substitution with another entity "of the same kind" rather than with something that is expected to be found in a scenario that is typical for the negated entity. In this sense, we might expect to find more

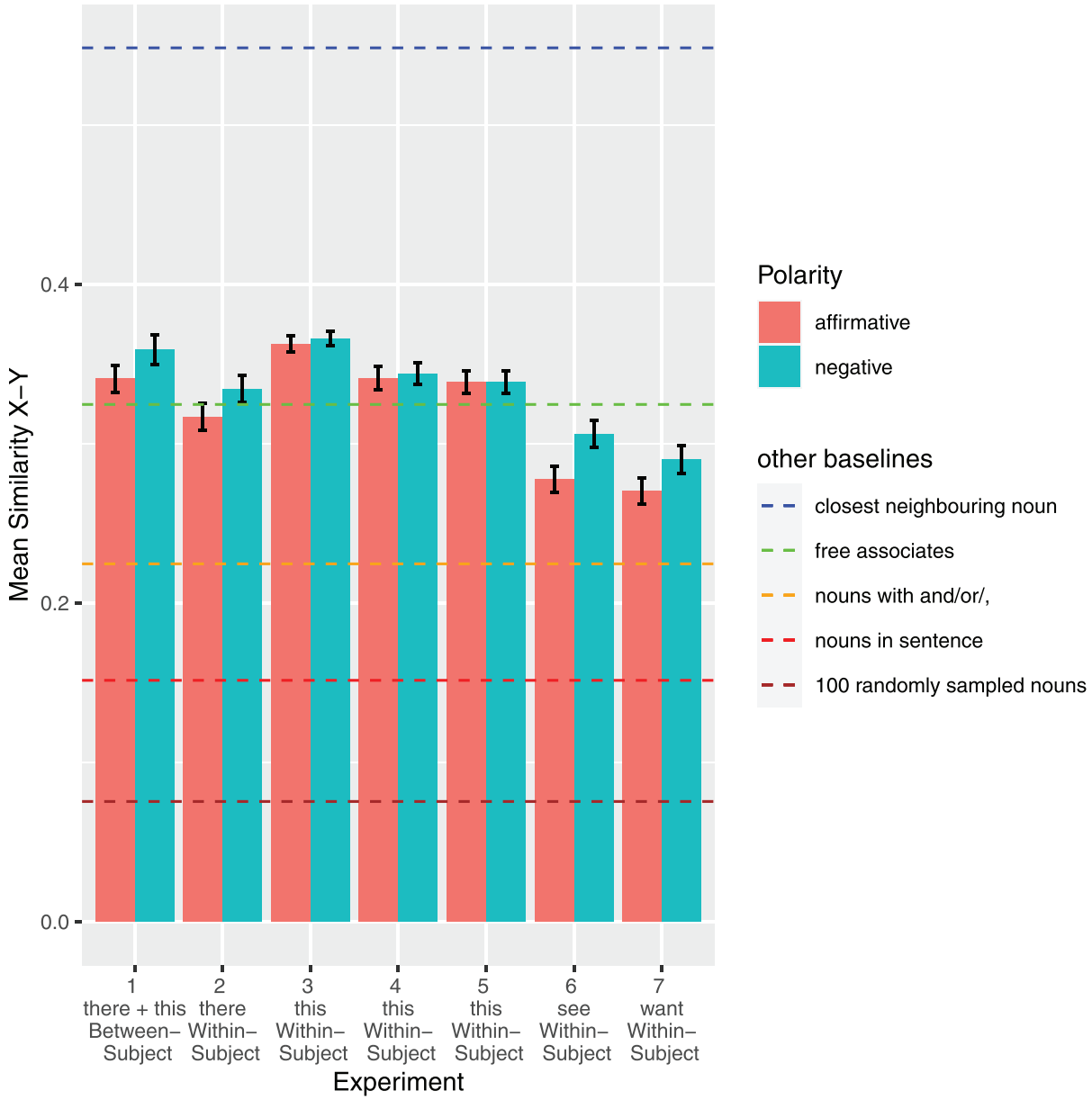


Fig. 8. General baselines and mean cosine similarity scores (aggregated by subject) of affirmative and negative conditions for each experiment reported in this paper. Error bars represent $\pm SE$ of the means.

co-hyponyms—whose relation is defined by sharing a close hypernym—in response to negative sentences than to affirmative sentences. A closer look at the data further supports this intuition: completions were classified as co-hyponyms when within WordNet (Miller, 1998) one of their noun synsets was found to share one "immediate" hypernym with a noun synset of the negated noun; the percentage of co-hyponyms produced in the negative condition is numerically higher than in the affirmative condition for all the within-subject design experiments, except for Experiment 4 (*This* context), where the pattern is reversed (see Table 1 for the percentages).

Table 1
Percentage of co-hyponyms per *Polarity* condition for Experiments 2–7

Experiment	*Sentence Type*	*Polarity*	Sentence	% Co-Hyponyms
2	there	Negative	There is no X here, but there is Y.	31
		Affirmative	There is X here, and there is Y there.	26
3	this	Negative	This is not X, it is Y.	37
		Affirmative	This is X, and that is Y.	35
4	this	Negative	This is not X here, but it is Y.	31
		Affirmative	This is X here, and that is Y there.	32
5	this	Negative	This is not X here, it is Y.	32
		Affirmative	This is X here, and that is Y there.	31
6	see	Negative	(Pron) see(s) no X, but (Pron) see(s) Y.	27
		Affirmative	(Pron) see(s) X and (Pron) see(s) Y.	20
7	want	Negative	(Pron) want(s) no X, but (Pron) want(s) Y.	22
		Affirmative	(Pron) want(s) X and (Pron) want(s) Y.	17

Similarly, an analysis employing WordNet path-length based similarity scores shows an advantage of negation (always significant, except for Experiment 5). In the analysis, we compared the model

$$Path \cdot Length \cdot Sim \sim Polarity + (1|Item) + (1 + Polarity|Subject) \tag{8}$$

against the baseline model

$$Path \cdot Length \cdot Sim \sim 1 + (1|Item) + (1 + Polarity|Subject). \tag{9}$$

In WordNet, path-length based similarity is the inverse of the length of the shortest path between two concepts (Jurafsky & Martin, 2018). For each negated noun/completion pair, we calculated the path-length based similarity of their two closest senses. This similarity measure can be expected to boost paradigmatic relations, as WordNet organizes word senses on a "is-a" hierarchy, and therefore might provide an additional hint on relational differences. Results are displayed in Table 2.

These patterns additionally assist the idea that the affirmative *This* contexts might be functionally more similar to their negative counterparts. In fact, completions in the *This* tasks display the highest degrees of overlap between the two conditions. Table 3 reports approximate measures of overlap (number of answers shared between affirmative and negative condition/total number of answers) as well as an approximate measure of variety (number of unique answers/total number of answers).

Given the possible prevalence of completions in paradigmatic relations in the case of negative sentences, one could wonder whether the difference in similarity scores is actually driven by the peculiarity of the semantic model employed. "Window"-type models have been argued to harvest paradigmatic similarities (Sahlgren, 2008). In fact, they overestimate paradigmatic over syntagmatic relations when predicting association strengths in priming tasks (Jones et al., 2006). In this sense, the difference in similarity scores would indeed reflect

Table 2
Summary of the results of the analysis employing path-length based similarity scores for Experiments 2–7

Experiment	*Sentence Type*	*Polarity*	Sentence	*p*
2	there	Negative	There is no X here, but there is Y.	
		Affirmative	There is X here, and there is Y there.	<.001
3	this	Negative	This is not X, it is Y.	
		Affirmative	This is X, and that is Y.	<.05
4	this	Negative	This is not X here, but it is Y.	
		Affirmative	This is X here, and that is Y there.	<.01
5	this	Negative	This is not X here, it is Y.	
		Affirmative	This is X here, and that is Y there.	.13
6	see	Negative	(Pron) see(s) no X, but (Pron) see(s) Y.	
		Affirmative	(Pron) see(s) X and (Pron) see(s) Y.	<.001
7	want	Negative	(Pron) want(s) no X, but (Pron) want(s) Y.	
		Affirmative	(Pron) want(s) X and (Pron) want(s) Y.	<.001

Table 3
Variety per *Polarity* condition (number of unique answers/total number of answers) and "Overlap" (number of answers shared between affirmative and negative condition/total number of answers) for Experiments 2–7

Experiment	*Sentence Type*	*Polarity*	Sentence	Variety	Overlap
2	there	Negative	There is no X here, but there is Y.	0.32	
		Affirmative	There is X here, and there is Y there.	0.32	0.24
3	this	Negative	This is not X, it is Y.	0.25	
		Affirmative	This is X, and that is Y.	0.27	0.30
4	this	Negative	This is not X here, but it is Y.	0.26	
		Affirmative	This is X here, and that is Y there.	0.27	0.27
5	this	Negative	This is not X here, it is Y.	0.27	
		Affirmative	This is X here, and that is Y there.	0.27	0.27
6	see	Negative	(Pron) see(s) no X, but (Pron) see(s) Y.	0.35	
		Affirmative	(Pron) see(s) X and (Pron) see(s) Y.	0.34	0.21
7	want	Negative	(Pron) want(s) no X, but (Pron) want(s) Y.	0.31	
		Affirmative	(Pron) want(s) X and (Pron) want(s) Y.	0.32	0.23

the more restrictive behavior of negation in terms of relational selection, but not necessarily mirror a difference in association strengths as its found in speakers. We ran the same analysis with an LSA model (Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998), which, on the other hand, has been argued to overestimate syntagmatic over paradigmatic relations (Jones et al., 2006). Although we could have expected the effect to disappear if the relational difference were the main drive, we observed the same pattern of results as reported in the main analysis (see Supporting Information Appendix C). Replicating the pattern with an LSA model makes the results more robust on a semantic similarity basis.

11. Discussion

The behavior of conversational negation often does not overlap with that of negation as a logical operator: by rejecting a state of affairs, conversational negation often suggests only a very small fraction (plausible alternatives) of the logically warranted set (the complement) of alternative state of affairs. Plausible alternatives tend to be the ones that are substitutable with the negated state of affairs across a large number of contexts. In fact, Kruszewski et al. (2016) showed that the plausibility of alternatives fostered by negation is highly correlated with the cosine similarity scores between the negated noun and the "alternative noun." As the correlation might not be exclusive to negation, we introduced comparison baselines to show that negation indeed evokes alternatives that are particularly similar to the negated entity. The main comparison baseline was the similarity of nouns conjuncted by *and* in affirmative sentences, which can be expected to score very high. Additionally, we inspected how similarity scores between negated nouns and alternatives compare to average similarity scores of closest neighbors in the semantic space, free associates, nouns co-occurring within [Noun] [*and*/*or*/,] [optional: *the*/*a*] [Noun] constructions, nouns co-occurring within a sentence and randomly sampled nouns.

The question was investigated in production. Subjects completed a series of cloze tasks: in each task, they were asked to complete negative sentences suggesting alternativehood with respect to a negated noun (e.g., *There is no goat here, but there is*_____) and/or affirmative conjunctive sentences that were kept structurally as similar as possible to the corresponding negative ones (e.g., *There is a goat here, and there is*_____ *there*). We employed a range of minimally constraining sentential contexts. Cosine similarity scores were calculated between each negated noun/completion noun pair for negative sentences, and between each given noun/completion noun pair for affirmative sentences.

Overall, the average similarity scores derived from the tasks attest themselves around the level of free associates (see Fig. 1 and Supporting Information Appendix B), which seems sensitive, as the minimally constraining contexts might have rendered the tasks akin to a free association task. Free associates are the most accessible words given a cue word, where ease of retrieval seems to be affected both by associative knowledge and by aspects of meaning (Nelson et al., 2004). In fact, they are well above words co-occurring in the *and/or/*, contexts found in a corpus, which likely span a wide range of constraints.

As the tasks were run online, we can imagine that a more controlled setting might have generated even higher average scores: in that case, subjects might be less prone on producing totally semantically unrelated completions (e.g., from participants' answers: "There is no ferry here, but there is *clown*.", "This is broccoli, and that is *software*.", etc.). In fact, any answer that after the cleaning procedure consisted of a single word present in the semantic space was retained as acceptable. Whereas not ideal—since it drastically reduces the number of usable data points—an analysis of the data excluding answers that were produced only once to a specific noun across a sentential context is likely to partly overcome this issue. Such an analysis indeed produces higher average similarity scores, as well as the same pattern of results as the main analysis (see Supporting Information Appendix D).

Table 4
Summary of the results of the main analysis

Experiment	*Sentence Type*	Task	*Polarity*	Sentence	*p*
1	there	1a	Negative	There is no X here, but there is Y.	
		1b	Affirmative	There is X here, and there is Y there.	
	this	1a	Negative	This is not X, it is Y.	
		1b	Affirmative	This is X, and that is Y.	.15
2	there	2	Negative	There is no X here, but there is Y.	
			Affirmative	There is X here, and there is Y there.	<.001
3	this	3	Negative	This is not X, it is Y.	
			Affirmative	This is X, and that is Y.	.31
4	this	4	Negative	This is not X here, but it is Y.	
			Affirmative	This is X here, and that is Y there.	.17
5	this	5	Negative	This is not X here, it is Y.	
			Affirmative	This is X here, and that is Y there.	.49
6	see	6	Negative	(Pron) see(s) no X, but (Pron) see(s) Y.	
			Affirmative	(Pron) see(s) X and (Pron) see(s) Y.	<.001
7	want	7	Negative	(Pron) want(s) no X, but (Pron) want(s) Y.	
			Affirmative	(Pron) want(s) X and (Pron) want(s) Y.	<.001

Fundamentally, comparisons were run between negative and affirmative sentences. These all resulted in a significant superiority of similarity scores for the negative sentences, except for one sentential context: all the variations of the *This* context consistently produced a null effect. A summary of the results can be found in Table 4.

For three out of four sentential contexts, similarity scores from negative sentences lie significantly above their affirmative counterparts: on average, negation seems to indeed prompt very similar alternatives. The effects appear to be quite strong, which, together with the consistent results of Experiments 3–5, suggest the null effect to be specific to the *This* context. After Experiment 3, it was initially hypothesized that a ceiling effect might be responsible for the absence of a difference in the *This* contexts. The idea seems unlikely given the results of Experiments 4 and 5, where the similarity scores lie below those of Experiment 3, therefore leaving room for more similar alternatives to be produced in the negative condition. On the other hand, it is not the case that alternativehood prompting is exclusive to negation. Many of the pragmatic functions typically ascribed to negation can also be conveyed by affirmation (Giora, 2006). Contrastive negation is not the only means of expressing contrast: similarly, contrast can be conveyed by intonational patterns, cleft constructions and word order (Silvennoinen, 2019). In the case of the *This* context, the affirmative sentence might equally convey a corrective reading, leading to the production of a substitute state of affairs. Further investigation is needed to confirm that the *This* context is indeed perceived as conveying contrastive communicative intentions, but this is out of the scope of the current paper. A first step in this direction was taken in our post hoc analysis (see next paragraph), which aligned to our intuitions. Nevertheless, the contrastive function seems to be overall more peculiar to

negation. It is, in fact, widely acknowledged that one of the core pragmatic purposes of negation is that of correcting a false presupposition (Givón, 1978; Wason, 1965). A false presupposition seems more justified if there is a reason to confuse two states of affairs (presupposed vs. actual), therefore a minimal difference between the two renders the use of negation more felicitous.

In addition to the planned analysis, we conducted a post hoc analysis in order to investigate whether alternatives might be constrained on a relational level with respect to the negated entity. A stronger preference for paradigmatic relationships was expected in the case of negative sentences: if negation fosters contrast between similar entities, we should expect more completions that are substitutable to the negated entity—such as co-hyponyms—as opposed to affirmation, where completions do not need to be interchangeable with the given entity. Completions were recognized as co-hyponyms with the help of Wordnet. Indeed, the percentage of co-hyponyms detected among the completions is higher for negative sentences than for affirmative sentences in five of six experiments. The speculation is additionally supported by higher average path-length similarity scores—a similarity measure that can be expected to boost paradigmatic relationships—in the negative conditions for most tasks. It is unclear though—and in need of further clarification—whether the paradigmatic/syntagmatic distinction is actually at work or the preference can be reduced to similarity alone. Finally, supporting our intuition that the affirmative *This* context might in fact behave the most similar to its negative counterpart (i.e., conveying contrast and suggesting alternatives), these tasks displayed the smallest difference in the percentage of co-hyponyms detected and the highest degree of overlap in the answers between affirmative and negative sentences.

Our results confirm the idea that linguistic negation acts very differently from a logical operator, proving itself to be in fact highly restrictive in the suggestion of alternative scenarios: naturally produced alternatives are very similar to the negated entity. Furthermore, the restriction appears to be even tighter than in the case of affirmation, where entities are expected to be limited by the probability of co-occurring in the same scenario. Although the logical use of negation would be largely permissive in the selection of plausible alternatives, conversational negation thus acts even more restrictively than conjunction within an affirmative context.

Acknowledgments

Open access funding enabled and organized by Projekt DEAL.

References

Baroni, M., Dinu, G., & Kruszewski, G. (2014). Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (Volume *1*: Long Papers, pp. 238–247). Stroudsburg, PA: Association for Computational Linguistics.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Blakemore, D. (1989). Denial and contrast: A relevance theoretic analysis of "but. *Linguistics and Philosophy*, *12*, 15–37.

Bruni, E., Tran, N.-K., & Baroni, M. (2014). Multimodal distributional semantics. *Journal of Artificial Intelligence Research*, *49*, 1–47.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of Cognition*, *1*(1), 9.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. New York: Harcourt Brace Jovanovich.

Cornish, E. R. (1971). Pragmatic aspects of negation in sentence evaluation and completion tasks. *British Journal of Psychology*, *62*(4), 505–511.

De Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*(1), 1–12.

De Saussure, F. (2011). *Course in general linguistics*. New York: Columbia University Press.

De Villiers, J. G., & Flusberg, H. B. T. (1975). Some facts one simply cannot deny. *Journal of Child Language*, *2*(2), 279–286.

Deutsch, R., Gawronski, B., & Strack, F. (2006). At the boundaries of automaticity: negation as reflective operation. *Journal of Personality and Social Psychology*, *91*(3), 385.

Dudschig, C., Mackenzie, I. G., Maienborn, C., Kaup, B., & Leuthold, H. (2019). Negation and the n400: Investigating temporal aspects of negation integration using semantic and world-knowledge violations. *Language, Cognition and Neuroscience*, *34*(3), 309–319.

Fischler, I., Bloom, P. A., Childers, D. G., Roucos, S. E., & Perry Jr, N. W. (1983). Brain potentials related to stages of sentence verification. *Psychophysiology*, *20*(4), 400–409.

Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory & Cognition*, *28*(7), 1109–1115.

Giora, R. (2006). Anything negatives can do affirmatives can do just as well, except for some metaphors. *Journal of Pragmatics*, *38*(7), 981–1014.

Givón, T. (1978). Negation in language: Pragmatics, function, ontology. In P. Cole (Ed.), *Pragmatics* (pp. 69–112). Leiden: Brill.

Glenberg, A. M., Robertson, D. A., Jansen, J. L., & Johnson-Glenberg, M. C. (1999). Not propositions. *Cognitive Systems Research*, *1*(1), 19–33.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan. (Eds.) *Speech acts* (pp. 41–58). Leiden: Brill.

Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun—An R package for computations based on latent semantic analysis. *Behavior Research Methods*, *47*(4), 930–944.

Günther, F., Dudschig, C., & Kaup, B. (2016). Latent semantic analysis cosines as a cognitive similarity measure: Evidence from priming studies. *Quarterly Journal of Experimental Psychology*, *69*(4), 626–653.

Günther, F., Rinaldi, L., & Marelli, M. (2019). Vector-space models of semantic representation from a cognitive perspective: A discussion of common misconceptions. *Perspectives on Psychological Science*, *14*(6), 1006–1033.

Harris, Z. S. (1954). Distributional structure. *Word*, *10*(2–3), 146–162.

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

Johnson-Laird, P. N., & Tridgell, J. (1972). When negation is easier than affirmation. *Quarterly Journal of Experimental Psychology*, *24*(1), 87–91.

Jones, M. N., Kintsch, W., Mewhort, D. J. (2006). High-dimensional semantic space accounts of priming. *Journal of Memory and Language*, *55*(4), 534–552.

Jurafsky, D., & Martin, J. H. (2018). Speech and language processing (draft). Available at: https://web.stanford.edu/~jurafsky/slp3/ed3book.pdf

Kaup, B., & Dudschig, C. (2020). Understanding negation: Issues in the processing of negation. In V. Déprez & M. T. Espinal (Eds.), *The Oxford handbook of negation* (pp. 635–655). Oxford, England: Oxford University Press.

Kruszewski, G., Paperno, D., Bernardi, R., & Baroni, M. (2016). There is no logical negation here, but there are alternatives: Modeling conversational negation with distributional semantics. *Computational Linguistics*, *42*(4), 637–660.

Landauer, T. K., & Dumais, S. T. (1997). A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*(2), 211.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284.

Lang, E. (2008). 26. koordinierende konjunktionen (Coordinative conjunctions). In A. von Stechow & D. Wunderlich (Eds.), *Semantik/Semantics* (pp. 597–623). Berlin: De Gruyter.

Lenci, A. (2008). Distributional semantics in linguistic and cognitive research. *Italian Journal of Linguistics*, *20*(1), 1–31.

Lenci, A. (2018). Distributional models of word meaning. *Annual Review of Linguistics*, *4*, 151–171.

Lenci, A., & Benotto, G. (2012). Identifying hypernyms in distributional semantic spaces. In **SEM 2012: The First Joint Conference on Lexical and Computational Semantics—Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)* (pp. 75–79). Stroudsburg, PA: Association for Computational Linguistics.

Lüdtke, J., & Kaup, B. (2006). Context effects when reading negative and affirmative sentences. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (Vol. *27*, pp. 1735–1740). Mahwah, NJ: Lawrence Erlbaum.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, *28*(2), 203–208.

Malchukov, A. L. (2004). Towards a semantic typology of adversative and contrast marking. *Journal of Semantics*, *21*(2), 177–198.

Mauri, C., & Van der Auwera, J. (2012). Connectives. In K. Allan & K. M. Jaszczolt (Eds.), *The Cambridge handbook of negation* (pp. 377–401). Cambridge: Cambridge University Press.

Miestamo, M. (2007). Negation—An overview of typological research. *Language and Linguistics Compass*, *1*(5), 552–570.

Miller, G. A. (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Moeschler, J. (1992). The pragmatic aspects of linguistic negation: Speech act, argumentation and pragmatic inference. *Argumentation*, *6*(1), 51–76.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers*, *36*(3), 402–407.

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.

Nordmeyer, A. E., & Frank, M. C. (2015). Negation is only hard to process when it is pragmatically infelicitous. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, Pasadena, CA (pp. 23–25).

R Core Team (2017). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sahlgren, M. (2008). The distributional hypothesis. *Italian Journal of Disability Studies*, *20*, 33–53.

Silvennoinen, O. O. (2019). *Contrastive negation: Constructional variation within and across languages* [Dissertation]. Helsinki: University of Helsinki.

Wason, P. C. (1961). Response to affirmative and negative binary statements. *British Journal of Psychology*, *52*(2), 133–142.

Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, *4*(1), 7–11.

Westfall, J., Kenny, D. A., & Judd, C. M. (2014). Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. *Journal of Experimental Psychology: General*, *143*(5), 2020.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supporting Information

Appendix B

Study 2: Capuano et al. (2023)

Memory & Cognition
https://doi.org/10.3758/s13421-023-01434-2

Activation levels of plausible alternatives in conversational negation

Francesca Capuano[1] · **Theresa Sorg**[1] · **Barbara Kaup**[1]

Accepted: 17 May 2023

Abstract
Negation is often used to contradict or correct (e.g. *There is no dog here.*). While rejecting some state of affairs that is presumed to hold for the recipient (e.g. *There is a dog here.*), the speaker might implicitly suggest a set of plausible alternatives (e.g. *There is a wolf instead.*). Prior work indicates that alternatives are highly relevant to the comprehension of sentences involving focus: in priming studies, listeners infer plausible alternatives to focused items even when they are not contextually available. So far it is unclear whether negation similarly activates an automatic search for plausible alternatives. The current study was designed to investigate this question, by looking at the activation levels of nouns after negative and affirmative sentences. In a series of priming experiments, subjects were presented with negative and affirmative sentences (e.g. *There is an/no apple.*), followed by a lexical decision task with targets including plausible alternatives (e.g. *pear*), as well as semantically related but implausible alternatives (e.g. *seed*). An interaction of Sentence Polarity and Prime-Target Relation was expected, with negation facilitating responses to plausible alternatives. Results of the first experiment were numerically in line with the hypothesis but the interaction just missed significance level. A post hoc analysis revealed the expected significant interaction. Possible roles of sentential context and goodness of alternatives are discussed. A further experiment confirms that the goodness of alternatives is in fact critical in modulating the effect.

Keywords Negation · Pragmatics · Alternatives · Activation levels · Priming

It has been long pointed out that one of the primary functions of negation in natural language is to contradict or correct (Clark and Clark, 1977; Givón, 1978; Wason, 1965; Kaup, 2009; Horn, 1989). The use of negation often presupposes the understanding of the speaker that the listener might erroneously believe the negated state of affairs to hold. Schindele et al. (2008) showed that Theory of Mind processes, i.e. the ability to put oneself in the mental state of the other person, are indeed necessary to understand the pragmatic aspects of negation. In Kruszewski et al. (2016) speakers judged a sentence like *It's not a dog, it's a wolf* as more plausible than a sentence like *It's not a dog, it's a screwdriver*. More specifically, they show a correlation to exist between the semantic similarity of the negated entity (*dog*) and its alternatives (*wolf* vs. *screwdriver*), and the plausibility judgments of the sentences: the more similar the two entities (e.g. *dog* and *wolf*), the more plausible was the sentence rated. Possibly, higher similarity corresponds to higher confusability between the two entities, therefore licensing the assumption of the speaker on the false presupposition held by the listener.

Capuano et al. (2021) confirmed that the preference for highly similar alternatives is indeed specific to negation, going beyond a general preference for semantically similar nouns within the same sentence. In a series of cloze tasks, they collected completions to four different minimal sentential contexts, both in the negative and in the affirmative form (e.g. *There is no dog here, but there is* _____ vs. *There is a dog here, and there is a* _____ *there.*)). In three out of four contexts, completions to the negative sentences appeared to be significantly more similar to the first-mentioned entity than in the case of the affirmative sentences, indeed confirming this to be a peculiarity of negation. In this study subjects were explicitly prompted for an alternative. To our knowledge, it has not been investigated whether negation automatically activates a search for

Francesca Capuano
francesca.capuano@uni-tuebingen.de

Theresa Sorg
theresa.sorg@uni-tuebingen.de

Barbara Kaup
barbara.kaup@uni-tuebingen.de

[1] Department of Psychology, University of Tübingen, Schleichstraße 4, 72076 Tübingen, Germany

Published online: 05 July 2023

plausible alternatives in case no alternative were explicitly solicited.

Alternatives are key to the successful comprehension of certain types of utterances. Following the work of Rooth (1992) on alternative semantics, focused items evoke alternatives which are relevant to the interpretation of an utterance (e.g. *[Mary]$_F$ likes Sue* evokes the set of propositions of the form *x likes Sue*, whereas *Mary likes [Sue]$_F$* evokes the set of propositions of the form *Mary likes x*). Recent priming studies have tapped into the psychological reality of this semantic theory, showing that listeners do infer alternatives to focused items, even when (Yan and Calhoun 2019; Braun and Tagliapietra 2010; Husband and Ferreira 2016; for a more comprehensive overview, see Gotzner and Spalek 2019). Using cross-modal lexical decision tasks, Braun and Tagliapietra (2010) investigated semantic priming of sentences differing in intonation pattern and in the semantic relation between the final word and the target word in Dutch. When an utterance was contrastively accented (e.g. *In **Florida** he photographed a **flamingo***), contextual alternatives (e.g. *pelican*) were activated preferentially compared to a neutral intonation condition, but non-contrastive semantic associates (e.g. *pink*) were not. Husband and Ferreira (2016) explored the time course of the priming effect in English, finding evidence for an initial activation of both contrastive and non-contrastive associates, followed by the selective deactivation of non-contrastive associates in the case of contrastively accented sentences. Yan and Calhoun (2019) shows preferential priming of contextual alternatives in Mandarin Chinese when the focus is realised via prosodic prominence, but not when it is realised only syntactically.

Alternatives might be equally relevant to the comprehension of negation. In fact, whereas negation comprehension has been translated psychologically to the general activation of inhibitory mechanisms (de Vega et al., 2016; Beltrán et al., 2019; Liu et al., 2020; Beltrán et al., 2021), the corrective function often linguistically attributed to this construction could find a psychological correlate in the selection of plausible alternatives (either by enhanced activation of plausible alternatives or by inhibition of implausible ones), similarly to what happens with contrastive focus. In that case, we might expect plausible alternatives to be activated preferentially after negative rather than after affirmative statements. This would be in line with existing evidence on activation levels after negation processing. Orenes et al. (2022, Experiment 1), using the priming methodology, found that negation does not decrease the activation level of the mental representation of the negated entity. For example, they found that there was no significant difference in reading time between the target assertions *There were roses and there were lilies* after reading a negated conjunction (*It is not the case that there are roses and there are lilies*) and after reading an affirmative conditional (*If there are roses, then there are lilies*). MacDonald and Just (1989), whilst observing that negation did decrease the activation level of the negated entity (inhibition), on the other hand found that concepts related to the negated entity were not significantly less active than in the case of affirmation. Therefore, negation might (or not) inhibit the activation of the negated concept but no evidence is found for the inhibition of its associates. Crucially, MacDonald and Just made no differentiation between associates that are also plausible alternatives with respect to the negated entity (e.g. *wolf* in relation to *dog*), and associates that are not (e.g. *bone* in relation to *dog*). Therefore, their results do not preclude the possibility of a different pattern for plausible vs. implausible alternatives, and specifically a facilitation of the first with respect to the latter.

Additionally, there is evidence that - whether mediated by the representation of the negated state of affairs (two-simulation hypothesis, see Kaup and Zwaan (2003); Kaup et al. (2007, 2005)) or not (e.g. Mayo et al. 2004) - speakers eventually reach a representation of the actual state of affairs as a result of processing negative statements. Kaup et al. (2006) presented subjects with sentences involving contradictory predicates (e.g. *The door is (not) closed/open.*). The subjects were then presented with a picture and asked to name aloud the entity depicted. The entity was varied according to whether it corresponded to the described or to the negated state of affairs. According to the sentence *The door is closed*, a closed door would correspond to the actual state and an open door would not. On the other hand, according to the sentence *The door is not closed*, an open door would correspond to the actual state and a closed door would not. In order to investigate the temporal characteristics of the representational process, the delay at which the image was presented was varied. One half of the subjects was presented with the image after 750 ms, the other half only after 1500 ms. With a delay of 750 ms, an effect of agreement was found in relation to the actual state in affirmative sentences (e.g. a closed door after the sentence *The door is closed.*), but not in negative sentences (e.g. an open door after the sentence *The door is not closed.*). With a delay of 1500 ms, this effect was also present for negative sentences - the subjects had thus mentally represented the actual state at this point in time, while at an earlier point in time the state to be negated was represented.

A sentence like *The door is not closed* might likely be uttered assuming that the person being addressed thought that the door was closed, and the speaker wanted to make them aware that this was not the case. In this example, the person addressed can easily infer the alternative state (an open door). In fact, *open* and *closed* are contradictory predicates, representing the only two possible states. The actual state of affairs (the alternative) is therefore confined to one possibility here. However, if one looks at the previously mentioned sentence *This is not a dog*, it is harder to determine what the

actual state of affairs might be. Seen from the perspective of set theory, any member of the complement set of *dog* would be consistent with *not a dog*. As we have seen though, some entities are more likely alternatives than others: in the case of words that don't relate to a direct opposite, negation acts a graded similarity function that produces a probability distribution over a restricted set of alternatives (Kruszewski et al., 2016). Given this uncertainty, it is unclear whether listeners still activate likely alternatives in their mental representation.

Based on these considerations, the present study will investigate whether negation leads to the activation of plausible alternatives. For this purpose, we constructed affirmative and negative prime sentences, each involving a concrete entity (e.g. *There is an/no apple.*). The presentation of the sentences was followed by a lexical decision task. Targets could either constitute a plausible alternative (e.g. *pear*) or a semantically related but implausible alternative (e.g. *seed*) with respect to the entity in the prime sentence. Unrelated words were also presented as targets for the sake of a manipulation check: as per literature, we expected both plausible and implausible alternatives to be responded to faster than unrelated words. Non-words were also used as targets to complete the lexical decision task. The experiment was conducted in German. Our main hypothesis was that, assuming negation leads to a search for alternatives, the difference in RTs between negative and affirmative sentences should be smaller in the case of plausible alternatives, compared to semantically related but implausible alternatives. More specifically, an interaction effect of Sentence Polarity and Prime-Target Relation is expected, with negation facilitating responses to plausible alternatives. Secondarily, in line with the literature on negation resulting in increased processing times, a main effect of Polarity can be expected. Nevertheless, no specific prediction is made for the main effect of Polarity within each Prime-Target Relation level: when we say that we expect negation to facilitate responses to plausible alternatives, we do not mean that we necessarily expect negation to display faster reaction times than affirmation to plausible alternatives, because we cannot rule out an underlying main effect of Polarity. Similarly, slower reaction times for implausible alternatives might not signal deactivation in absolute terms because they could be confounded with the slowing effect of negation. What stays informative in light of these considerations is the interaction effect, because any pattern of results displaying a reduced effect of negation in the plausible alternatives condition is consistent with the hypothesis of their *facilitated* activation after negation, although an inhibition of implausible alternatives should equally not be ruled out. In this sense, *facilitation of plausible alternatives* is an umbrella expression for both possibilities.

Experiment 1

Participants

Data was collected until 240 usable participants were reached. The sample size was determined based on a previous version of this study which employed slightly different materials that later on turned out to be inadequate. This will be elaborated on in the General Discussion. In total, 294 subjects were tested (201 female, 89 male, 4 diverse) with an age range of 19 to 59 years ($mean = 24.52$, $sd = 7.60$). Subjects who stated that they were not native German speakers were excluded ($n = 8$) as well as subjects who stated to have already participated in a very similar study (e.g. the previous version of this study) ($n = 27$). The recruitment took place partly via the mail server of the University, and partly via Prolific (Palan and Schitter, 2018). All subjects gave written informed consent.

Materials

We created affirmative and negative prime sentences of the form *There is [a/an]/[not a/not an] X* (German *Dort ist [ein/e]/[kein/e] X*). Ninety-six common nouns were selected to replace *X*, once in the affirmative and once in the negative form. For each of the 96 nouns (e.g. *apple*), two target words were selected, varying in terms of the relationship they bear with the noun: one plausible alternative (e.g. *pear*) and one semantic associate that is an implausible alternative (e.g. *seed*). Additionally, one semantically unrelated noun (e.g. *brush*) was added for the sake of a sanity check. In fact, both plausible and implausible alternatives should be primed more than an unrelated word. In order to help us construct these items, we ran a cloze task prior to the main study. For this task we chose 100 concrete high frequency nouns (e.g. *apple*). One hundred subjects were instructed to complete sentences such as *This is not an apple, it's* _____ with either just a noun (e.g. *pear*) or an indefinite article plus a noun (e.g. *a pear*). Like the main experiment, the cloze task was run in German. The resulting cloze frequencies helped us construct the majority of the materials (77 out of 96 items) in that, where permitted by the length and frequency match requirements (see below), the plausible alternatives were selected among frequent cloze answers. The remaining items ($n = 19$) were crafted by the authors by intuition. The implausible alternatives were selected among nouns with high similarity that were not good cloze completions. We took care that the implausible alternatives were not cohyponyms, as the post hoc analysis in Capuano et al. (2021) suggests that cohyponyms are particularly good alternatives

to negated nouns, at least as far as minimal contexts are concerned.

We controlled for length (number of characters) and frequency (raw counts) between targets across target relations. The mean length for the *plausible alternatives* was 6.43 ($sd = 2.41$), 6.17 ($sd = 1.97$) for the *implausible alternatives* and 6.36 ($sd = 1.83$) for the *unrelated* targets. Plausible alternatives did not differ from implausible alternatives ($t(95) = 0.91$, $p = .37$), nor from unrelated targets ($t(95) = 0.20$, $p = .84$). Implausible alternatives and unrelated targets were also paired ($t(95) = -0.69$, $p = .49$).

Target frequency counts were extracted from the deWaC corpus (Baroni et al., 2009) and employed for the matching procedure. The mean frequency for the plausible alternatives was 20906 ($sd = 56102$), 24946 ($sd = 41120$) for implausible alternatives and 27328 ($sd = 95932$) for unrelated targets. All the pairs of conditions were matched: plausible and implausible alternatives ($t(95) = -0.56$, $p = .58$), plausible alternatives and unrelated ($t(95) = -0.57$, $p = .57$), implausible alternatives and unrelated ($t(95) = -0.22$, $p = .83$).

Additionally, cosine similarity scores were calculated for each pair of noun in the sentence (X) and target, employing the LSAfun package with the dewak100k lsa Wordspace (Günther et al., 2015). Plausible alternatives were significantly more similar to the noun ($mean = 0.67$, $sd = 0.20$) not only with respect to unrelated targets ($mean = 0.21$, $sd = 0.14$; $t(95) = 18.15$, $p < .001$), but also to implausible alternatives ($mean = 0.48$, $sd = 0.21$; $t(95) = 6.22$, $p < .001$). Implausible alternatives were more similar to the noun than unrelated targets ($t(95) = 11.61$, $p < .001$). A match between plausible and implausible alternatives could not easily be achieved, as cohyponyms (plausible alternatives) normally tend to score higher on similarity scores than nouns in other semantic relationships (implausible alternatives). Although desirable to achieve an even cleaner design, the match between plausible and implausible alternatives is not needed for the testing of our main hypothesis, since we are testing for the interaction of Polarity and Relation, not for the main effect of Relation.

The two polarity levels of the sentence (*affirmative* vs. *negative*) and the three relation types between the noun in the sentence and the target (*plausible alternative* vs. *implausible alternative* vs. *unrelated*) resulted in six experimental conditions. The conditions were counterbalanced across participants, resulting in a total of six experimental lists with 96 experimental items each. In the experimental material, some words appeared as targets to more than one item. It was taken care that no target would appear more than once in any single list. Additionally, 96 filler sentences were created. These were all in the form of the experimental sentences (e.g. *There is [a/an]/[not a/not an] X*). The target words for the fillers were non-words that were created with the help of the pseudo-word generator Wuggy (Keuleers and Brysbaert, 2010) from the experimental targets of the corresponding list. Each experimental list thus contained 96 experimental sentences (48 affirmative, 48 negative) with 96 target words (32 for each type of relation) and 96 filler sentences with 96 non-words as targets.

In order to ensure that the participants had read the sentences and, above all, had processed the negation, they were prompted to re-type the previously read sentence for 48 of the 192 trials. Half of these were experimental sentences and the other half were filler sentences. Forty usable subjects were collected for each list. The complete collection of items, together with the collected data and analyses scripts of all the experiments presented in this paper can be found at https://osf.io/p5g8u/?view_only=3f40dc06bfa84ac09ce66120c6e71ae4. The experiment was programmed using jsPsych (De Leeuw, 2015), a JavaScript library that can be used to create online experiments.

Procedure

The task of the participants was to read each sentence presented on the screen, then judge whether a target string presented thereafter was either an existing word or a non-word (lexical decision task). At the beginning of each trial, the word *Attention*[1] appeared in red for 500 ms. A white screen was then shown for 200 ms, followed by the sentence, presented in its entirety (e.g. *There is no apple*). The participants could read the sentence at their own pace, then press the space bar to proceed. According to Kaup et al. (2006), speakers arrive at the factual representation of negative sentences some time between 750 and 1500 ms after sentence processing. We therefore start off with a delay of 1000 ms. After pressing the space bar, a fixation cross was presented for 1000 ms, then the target string appeared (e.g. *pear*). Participants were instructed to react as quickly as possible by pressing either one of two buttons: *k* for words and *d* for non-words. If the target was not responded to within 3500 ms, the message *Too slow! Please react faster!* was displayed and the experiment would proceed to the next trial; otherwise, feedback was provided on the correctness of the lexical decision. After the lexical decision task, subjects were occasionally prompted to re-type the previously read sentence in an input field. The initial instructions were followed by 10 practice trials. Then, the randomized 192 trials from the corresponding experimental list started. At the end of the experiment subjects were requested to provide their age, gender and handedness.

[1] For simplicity and for the sake of the procedure explanation, material is translated to English.

Results

Subjects who failed to retype at least 36 of the 48 sentences were excluded from the analysis ($n = 14$). Individual trials were excluded if the lexical decision was incorrect, too slow (>3500ms) or too fast (<200 ms). Only subjects with at least 154 (80 %) correct lexical decisions were included in the analysis ($n = 1$ subject excluded). Trials in which the reading time for the sentence was too short (<350ms) were excluded. Lexical decision RTs deviating more than 2.5 standard deviations from the mean of the corresponding condition (Polarity x Relation x Subject) were also excluded. Subjects who were left with less than eight data points per condition after the cleaning procedure were eliminated ($n = 4$). 18,37% of the initial subjects' datasets was excluded.

The data were analyzed with linear mixed effect models using the *lme4* package in *R* (Bates, 2005). In order to run a sanity check and ensure that both plausible and implausible alternatives were activated more strongly than unrelated, Model 1 was fit to the data *unrelated* trials

$$rt \sim Relation + (1|Item) + (1|Subject) \quad (1)$$

and compared against the baseline Model 2

$$rt \sim 1 + (1|Item) + (1|Subject) \quad (2)$$

through a Likelihood-Ratio Test (LRT). Treatment coding was employed, with *unrelated* as reference level. Model 1 explained the data significantly better than Model 2 ($\chi^2(2) = 102.56$, $p < .001$). Both the plausible and implausible alternatives differed significantly from the unrelated words ($\beta = -18.49$, $p < .001$ and $\beta = -11.25$, $p < .001$ respectively). On the other hand, a model with Polarity as fixed effect did not provide any improvement over Model 2 ($\chi^2(1) = 0.21$, $p = .65$).

In order to test for the interaction of Relation and Polarity, unrelated words were omitted. A null model (Model 3), which included the two fixed factors Relation and Polarity, as well as items and subjects as random factors, was compared to Model 4, additionally including the interaction effect. Default treatment coding was employed, with *aff:plausible alternative* as reference level. These models and this analysis were preregistered after they were employed on the pilot data (https://osf.io/7qxne?mode=&revisionId=&view_only=) and are then used across experiments for consistency.[2]

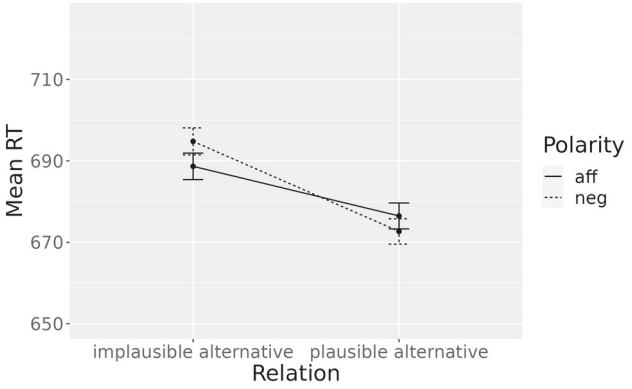


Fig. 1 **Experiment 1** - Mean RTs. Error bars represent ± se of the means

$$rt \sim Relation + Polarity + (1|Item) + (1|Subject) \quad (3)$$

$$rt \sim Relation * Polarity + (1|Item) + (1|Subject) \quad (4)$$

Models 3 and 4 were fit to the experimental data. The interaction just missed the significance level ($\chi^2(1) = 3.24$, $p = .07$). The mean RTs per condition are shown in Fig. 1. Model 4's estimates for the fixed effects across all the experiments reported in this paper can be found in Table 1.

The pattern of means consistently adheres to the hypothesis, and the interaction just misses the significance level. A consideration to make is that not all plausible alternatives were *equally good* alternatives: some had higher cloze frequencies than others, and for 19 items cloze frequencies are not available as they were crafted by intuition. We can expect the plausible alternatives with the highest cloze frequencies to be more likely candidates for an enhanced activation, as a larger proportion of subjects can be expected to prefer them. This possibility was explored in a post hoc analysis.

Post hoc analysis

In order to explore the role of alternative *goodness*, we analysed the data of Experiment 1 whilst retaining only the items with a cloze frequency above the median (>28).

[2] As LRT was shown to produce anti-conservative p values, especially with few observations (Kuznetsova et al., 2017), we also report the results of the analysis with the *lmerTest* outputs in the Appendix (Table 2), which do not diverge at all from our main analysis. As requested by a reviewer, we also report the results of the maximal models that we could fit for each experiment without encountering convergence issues. See Table 3 for the results of the maximal model fit on Experiment 1. Similarly, these do not diverge from the main analysis.

Table 1 **Model 4.** Fixed effects estimates

	Experiment 1	Experiment 2
(Intercept)	656.45	679.68
Polarityneg	0.52	0.33
Relationimplausible alternative	16.36	23.66
Polarityneg:Relationimplausible alternative	16.90	12.62

A main effect of Relation showed ($\chi^2(2) = 68.64$, $p < .001$): plausible and implausible alternatives differed significantly from the unrelated words ($\beta = -32.58$, $p < .001$ and $\beta = -9.00$, $p < .05$ respectively). No effect of Polarity was observed ($\chi^2(1) = 2.36$, $p = .12$), but there was a significant Relation x Polarity interaction ($\chi^2(1) = 4.23$, $p < .05$) in the expected direction: plausible alternatives were facilitated after negation with respect to implausible alternatives, compared to the pattern of activation after affirmatives. The *lmerTest* ANOVA table for the interaction model is in the Appendix (Table 4). The mean RTs per condition are shown in Fig. 2. In conclusion, the goodness of alternatives seems to be an influential factor driving the effect.[3]

Experiment 2

Based on the results of the post hoc analysis of Experiment 1 on the *best* items, we ran a second Experiment to investigate the expected interaction on a new sample. For that, we first conducted a power analysis on Experiment 1 selecting only the 48 best items in terms of the cloze frequencies obtained in our cloze task. The power analysis with 1000 simulations ($\alpha = .05$) resulted in 85.8% power when employing 540 subjects.

Participants

Data was collected until 540 usable participants were reached. In total, 577 subjects were tested (382 female, 184 male, 11 diverse) with an age range of 18 to 59 years ($mean = 26.40$, $sd = 8.48$). Subjects who stated that they were not native German speakers ($n = 4$), as well as subjects who stated to have already participated in a very similar study (e.g. Experiment 1 or a previous version) ($n = 20$) were excluded from the analysis. The recruitment took place via Prolific (Palan and Schitter, 2018). All subjects gave written informed consent.

Materials

We used the 48 items of Experiment 1 (i.e. half) with the highest cloze frequencies. Fillers ($n = 48$) and target words were created analogously to the previous experiments. The mean length for the *plausible alternatives* was 6.33 ($sd = 2.39$), 6.31 ($sd = 1.84$) for the *implausible alternatives* and 6.27 ($sd = 2.03$) for the *unrelated* targets. Plausible alternatives did not differ from implausible alternatives ($t(47) = 0.06$, $p = .95$), nor from unrelated targets ($t(47) = 0.14$, $p = .89$). Implausible alternatives and unrelated targets were also paired ($t(47) = 0.10$, $p = .92$). The mean frequency for the alternatives was 29047 ($sd = 75968$), 24070 ($sd = 39998$) for related and 42752 ($sd = 132151$) for unrelated targets. Again, all the pairs of conditions were matched: plausible and implausible alternatives ($t(47) = 0.40$, $p = .69$), plausible alternatives and unrelated ($t(47) = -0.62$, $p = .54$), implausible alternatives and unrelated ($t(47) = -0.92$, $p = .36$). Plausible alternatives were significantly more similar to the noun ($mean = 0.70$, $sd = 0.18$) not only with respect to unrelated targets ($mean = 0.21$, $sd = 0.15$; $t(47) = 14$, $p < .001$), but also to implausible alternatives ($mean = 0.47$, $sd = 0.20$; $t(47) = 5.8$, $p < .001$). Implausible alternatives were more similar to the noun than unrelated targets ($t(47) = 9.2$, $p < .001$).

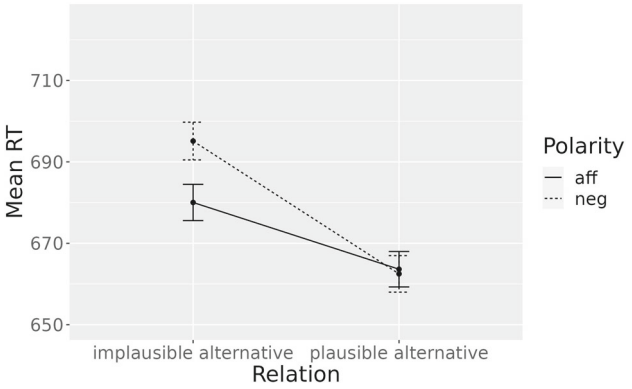


Fig. 2 **Post hoc Analysis** - Mean RTs when retaining only the items with the highest cloze frequencies. Error bars represent ± se of the means

[3] As requested by a reviewer, we additionally ran an analysis with Cloze Percentage as a covariate. The model is in Table 5 in the Appendix. In that model the three-way interaction with Cloze Percentage is not significant, but this does not invalidate our exploratory analysis, because the goodness of alternatives might not be a strictly continuous measure. Moreover, the intuition of the post hoc analysis was tested in a completely new experiment, erasing the possibility of any posthoc manipulation.

Procedure

The procedure was the same as in Experiment 1.

Results

The data analysis procedure was the same as in Experiment 1. Nine subjects failed to retype at least 18 of the 24 to-re-type sentences. Three subjects did not satisfy the minimum 80% accuracy criterion in the lexical decision task. Finally, four subjects were left with less than four observations in at least one experimental condition and were therefore excluded. In total, 6,64% of the collected subjects' datasets was excluded.

Model 1 explained the data significantly better than Model 2 ($\chi^2(2) = 126.49$, $p < .001$). In contrast to the prior experiment, the plausible alternatives - but not the implausible alternatives - differed significantly from the unrelated words ($\beta = -32.45$, $p < .001$ and $\beta = -2.67$, $p = .41$, respectively). This is maybe attributable to the fact that the best items that we selected for Experiment 2 by chance have particularly high mean frequency in the unrelated condition. Although the t-test shows no significant difference in the means, this might be driven by the larger standard deviations and reduced degrees of freedom. As in the prior experiment, a model with Polarity as fixed effect did not provide any improvement over Model 2 ($\chi^2(1) = 1.20$, $p = .27$). Models 3 and 4 were again fit to the experimental data. This time, the interaction reached the significance level ($\chi^2(1) = 3.87$, $p < .05$). There was a significant Relation x Polarity interaction in the expected direction with plausible alternatives being more facilitated after negation with respect to implausible alternatives, compared to the pattern of activation after affirmatives. The mean RTs per condition are shown in Fig. 3,[4]

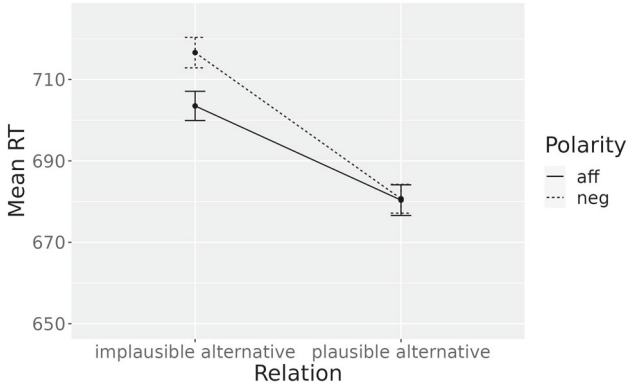


Fig. 3 **Experiment 2** - Mean RTs when retaining only the items with the highest cloze frequencies. Error bars represent ±se of the means

Internal meta-analysis

We ran a random-effects meta-analysis with the R package *metafor* (Viechtbauer, 2010) to determine the reliability of our effect across experiments, including the pilot and the discarded study ("Experiment 0"). The parameter coefficient for the interaction of Polarity and Prime-Target Relation is significant ($\beta = 10.33$, $95\%CI[3.83, 16.84]$, $se = 3.32$, $p < .01$). Figure 4 shows the forest plot of the meta-analysis.

[4] The analyses with *lmerTest* with the maximal converging model and the model with the cloze probabilities are in the Appendix (Tables 6, 7 and 8). The results do not diverge from the main analyses nor from the considerations made on Experiment 1, except for the main effect Polarity, which reaches significance with *lmerTest*.

Discussion and conclusions

Overview of the study

In natural language, negation often serves the purpose of correcting a previously held assumption. In this sense, it might prompt a search for a plausibly *correct* alternative. A correction implies some inaccuracy that is more justified, the more it is confusable with the rectification (≈ the more it bears resemblance to it). In fact, it was shown that, out of context, plausible alternatives to negated nouns tend to be very similar to the noun, often exceeding the similarity between nouns connected by *and* within an affirmative sentence.

The current study was designed to investigate whether negation facilitates the activation of plausible alternatives with respect to affirmation. It has been demonstrated that the comprehension of linguistic constructions that are semantically represented through alternatives sets - such as sentences involving focused items - are reflected in actual psychological mechanisms of activation of contextually plausible alternatives (Braun and Tagliapietra, 2010; Yan and Calhoun, 2019; Husband and Ferreira, 2016). Negation processing, instead, has been only generally linked to inhibitory mechanisms (Beltrán et al., 2019, 2021; de Vega et al., 2016; Liu et al., 2020). Nonetheless, its corrective connotation gives reason to question whether it equally generates a search for plausible alternatives. For this sake, we designed a priming experiment where minimalistic sentences varying in polarity (affirmative vs. negative) were used as primes, and nouns varying in their relationship with respect to the noun in the sentence (plausible vs. implausible alternative) were used as targets. An interaction effect of Sentence Polarity and Prime-Target Relation was expected between plausible and implausible alternatives, with negation facilitating the activation of plausible alternatives.

Our first experiment was based on a previous version that employed item materials of the form *This is [a/an]/[not a/not*

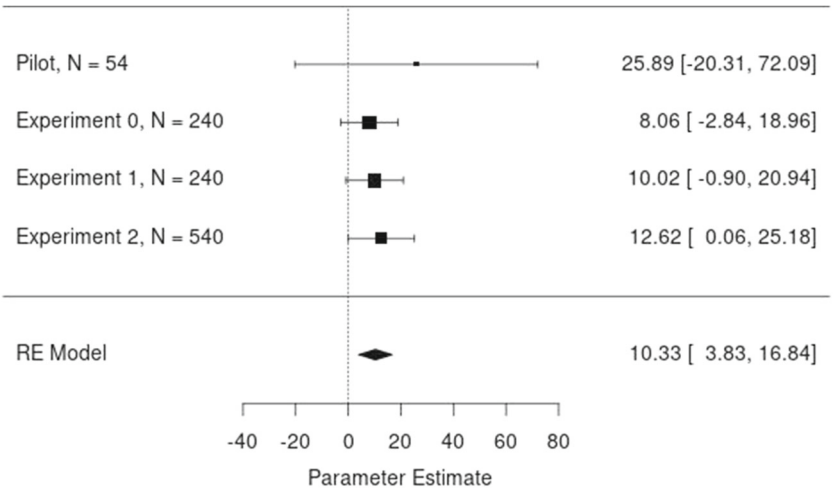


Fig. 4 Meta-Analysis - Forest plot of the meta-analysis on the interaction effect of Polarity and Prime-Target Relation

an] X (German *Das ist [ein/eine]/[kein/keine] X*). Using those sentential contexts, we had run a pilot study ($n = 60$) which did not provide evidence for the effect, but displayed a numeric tendency in the expected direction. Based on a power analysis of the pilot study, we collected data from 240 usable subjects. The item materials were identical to those used in Experiment 1, except for the different sentential context. This larger experiment replicated the results of the pilot (no interaction, but overall means numerically in line with the expectations). Alongside, Capuano et al. (2021) determined that the sentential contexts of the form *This is (no) X* are particularly infelicitous in demonstrating differences between affirmative and negative sentences. Specifically, we showed consistently no significant difference between affirmative sentences of the form *This is a goat, and that is a* _____ and negative sentences of the form *This is not a goat, it is a* _____ in terms of the similarity of the given noun to the noun they tend to be completed with. The same study though did find such a difference for three other sentential contexts (e.g. *There is no goat here, but there is a* _____ *there*). We concluded that the *This* sentential contexts might be a case where the affirmative version conveys a corrective reading just like the negative version, leading to the production of a substitute state of affairs. These preliminary studies were therefore discarded, but we still employed them to establish a reasonable sample size for Experiment 1, which was aimed at testing a sentential context that in Capuano et al. (2021) had displayed a significantly different behaviour depending on polarity (i.e. the *There* context). Whereas the interaction just misses the significance level ($p = .07$), the pattern of results of Experiment 1 is consistent with the hypothesis.

The question arose, whether the non-significant effect is due to non-unequivocal activation preferences. Differently from contradictory predicates, non-contradictory items can give rise to differential distributions of alternatives' activation. One can therefore expect some items to more uniformly give rise to the activation of a specific alternative across participants, i.e. the alternatives with the highest cloze probabilities. The plausible alternatives employed in our study differed in terms of cloze probability. In fact, an analysis retaining only the items with the highest cloze frequency alternatives (i.e. the *best* items) resulted in a significant interaction for Experiment 1. To collect additional evidence that the *goodness* of alternatives is the critical factor for the effect to emerge, we ran Experiment 2, based on a power analysis on the data for the *best* items of Experiment 1. Indeed, Experiment 2 produced the expected interaction, with the facilitation of *good* alternatives after negation.

Discussion of the results

Interaction effect of polarity and prime-target relation

Our results show that, when employing properly powered designs and appropriate items, a facilitative effect of negation to plausible alternatives can be detected, thereby confirming our hypothesis that negation can activate plausible alternatives also in the case of non-binary predicates.

As already clarified in the introduction, this facilitation is not to be understood in absolute terms, but rather with respect to what happens in the baseline condition (implausible alternatives). Even though at first sight the means suggest that negation inhibits implausible alternatives rather than facilitates plausible alternatives (because negation shows slower RTs than affirmation for implausible alternatives, but not for plausible alternatives), we should not forget that negation is commonly associated with longer processing times. A direct comparison of Polarity levels within each Prime-Target Relation level is therefore not very informative to our main hypothesis: the slower RTs in the case of implausible alternatives might result from a general slowing effect of

negation, and not from a specific inhibition of implausible alternatives.

Main effect of polarity

Contrary to our expectations, a main effect of sentence Polarity was never found. This could be due to our analysis being carried out exclusively on the reaction times of the lexical decision task. A slowing effect of negation might be visible only on sentence reading times - which were not analysed - without carrying over to the lexical decision task. Subjects determined themselves whether they were done reading the sentences before proceeding to the lexical decision task: by then, sentence processing might have been completed for both negative and affirmative sentences. Another possibility to consider is that precisely a facilitative effect of negation in the case of plausible alternatives might have wiped out a general Polarity effect.

Future directions

The finding of an interaction is particularly meaningful when we consider that our experimental items suffered from shortcomings due to the difficulty to control for multiple sources of variance. The minimalistic contexts, employed for comparability with Capuano et al. (2021) and for ease of collection of the alternatives, might have led the subjects to focus only on the noun and on the presence of the negation marker instead of reading the whole sentence. This might have rendered our paradigm more akin to a single word priming paradigm. In fact, the control task to retype the prime sentence did not ensure sentence-level comprehension but a more *semantic* task was difficult to devise with such minimalistic sentences. Both points potentially hindered the detection of the effect, which as a consequence becomes even more outstanding. Noticeably, the direction of the interaction in terms of trends remains consistent across all experiments, which makes it less likely that the findings are due to chance. This is further confirmed by the meta-analysis, which registers a significant overall interaction. Nevertheless, further effort should go into developing an alternative experimental design that can more neatly isolate the effect.

The use of minimalistic contexts also puts some limitations to the generalisability of the conclusions. Future investigation will need to extend the findings to other sentential contexts to make sure they are applicable to a general use of negation. The minimalistic contexts also circumscribed the types of alternatives investigated. In our study, plausible and implausible alternatives, apart from being distinguishable through differences in cloze task probabilities, stand systematically in different semantic relations with the prime noun: plausible alternatives are cohyponyms, whereas implausible alternatives stand in different relations to the prime. Similarly, Husband and Ferreira (2016)'s *contrastive associates* seem to be cohyponyms, whereas *non-contrastive associates* are associates in other kinds of semantic relations. The same goes for Yan and Calhoun (2019)'s materials. Braun and Tagliapietra (2010) go even further and sometimes employ different parts of speech such as advjectives as non-contrastive alternatives. We think that good alternatives are not limited to the relationship of cohyponymy, but by availability and how much overlap there is between the entities that is functional to the substitution in a specific context. Cohyponyms can substitute an entity in a wide range of contexts, so they tend to be the best alternatives when the contexts are not too restrictive. If negation activates a general search for plausible alternatives that are *contextually relevant*, we should be able to detect the same effect independently of semantic relation.

Finally, although we cannot exclude that the interaction effect might be driven by an inhibition in the activation of implausible alternatives rather than by the enhanced activation of plausible ones, Dennison and Schafer (2017) provides evidence in line with a progressive deactivation of less relevant alternatives, both in the case of contrast expressed through intonational form and in the case of contrast expressed through explicit negation. The study though is again confined to binary predicates. Further research in the time course of these activations after non-binary negation is needed to set apart the two processes more clearly (inhibition vs. enhanced activation).

Conclusive remarks

Alternatives in the context of negation most commonly refer to the contrast between the negated and the expressed proposition (Repp and Spalek, 2021); particular attention had been devoted to the time course of their access and integration in the mental model of the listener. Alternatives as the ones we refer to in the present study are traditionally investigated under the heading of *focus alternatives*, especially in relation to prosodically and syntactically marked focus. Different types of alternatives interact though, and disparate domains of alternatives have also been investigated jointly (Repp and Spalek, 2021). The current study suggests that negation functions as a (contrastive) focus marker, triggering focus alternatives without discourse context, and without explicit prosodic marking. In fact, the pragmatic functions attributed to negation resonate with the notion of contrast delineated by Zimmermann et al. (2008): contrastive focus expresses the speaker's assumption that the listener does not expect the upcoming information; as such, it signals the need for a shift in the interlocutor's assumptions and an update of their common ground. Therefore, relevant alternatives are not simply dictated by semantic similarity, but by speakers' expectations on the status of their common ground, whereby

semantic similarity is just a byproduct of the presentation of stimuli out of context. The discourse context is responsible for the restriction and therefore selection of the relevant alternatives.

Orenes et al. (2014) showed that, after hearing a sentence such as *The figure is not green*, subjects ended up fixating the alternative (e.g. a blue figure) whenever only two concurrent alternatives were offered by a visual world (a green and a blue figure), but stayed fixated on the green figure when more alternatives were presented (e.g. green, blue, yellow, pink). They conclude that alternatives are activated when there are only two, but not when there are more than two. Whereas no generalized experience suggests that *blue* is a better alternative to *green* than *yellow* though, there is reason to assume that this is not always the case whenever more than one alternative is available (e.g. some entities can be widely agreed upon to be *better alternatives* than others). As a more general criterion, our study suggests that the prominence (rather than the number) of potential alternatives might be the decisive factor determining the activation.

In conclusion, the evidence presented in this study supports the hypothesis that negation involves processing mechanisms that favour plausible alternatives. This seems the case even for non-binary negation, but might be confined to instances where the negated content displays particularly prominent alternatives. The psychological relevance of alternatives in the processing of negation is akin to the mechanisms demonstrated to be at play in the comprehension of structures marked with phonological focus. Therefore, the notion of contrastive focus might need to broaden to include negative constructions. The nature of the 'preference' for plausible alternatives in the case of negation is still unclear, potentially corresponding either to a preferential activation of plausible alternatives or to a selective deactivation of implausible ones. It is possible that the evidence for the use of inhibitory mechanisms in negation comprehension reflects exactly the process of deactivation ofimplausible alternatives. This issue needs further investigation and might benefit from the inspection of the time course of the candidates' activation.

Funding This work was supported by the Collaborative Research Centre 833 (SFB833) *The Construction of Meaning*/Z2 project appointed to Sigrid Beck and Barbara Kaup by the German Research Foundation (DFG).

Data Availability The materials, the data and the analyses scripts of all the experiments presented in this paper can be found at https://osf.io/p5g8u/?view_only=3f40dc06bfa84ac09ce66120c6e71ae4.

Declarations

Ethics approval This research was approved by the Ethics Committee for Psychological Research of the University of Tübingen.

Informed consent All subjects gave written informed consent to participation and anonymised data usage.

Conflicts of interest The Authors declare that there is no conflict of interest.

Appendix

Additional analyses

Table 2 **Experiment 1** - ANOVA table provided by the *lmerTest* package for the same analyses as in the main text

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Polarity	5370.65	5370.65	1.00	14139.72	0.19	0.6614
Relation	1095450.78	1095450.78	1.00	14144.51	39.13	0.0000
Polarity:Relation	90776.33	90776.33	1.00	14140.10	3.24	0.0718

$rt \sim Polarity * Relation + (1|Item) + (1|Subject)$

Table 3 **Experiment 1** - Maximal model reduced until convergence

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Polarity	5065.38	5065.38	1.00	235.79	0.19	0.6664
Relation	221686.98	221686.98	1.00	92.33	8.15	0.0053
Polarity:Relation	93512.24	93512.24	1.00	13822.81	3.44	0.0637

$rt \sim Polarity * Relation + (1 + Relation|Item) + (1 + Polarity|Subject)$

Table 4 Post hoc Analysis - Interaction model fit on the best items

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Polarity	90600.99	90600.99	1.00	6868.14	3.31	0.0690
Relation	1047570.66	1047570.66	1.00	6949.56	38.23	0.0000
Polarity:Relation	115858.85	115858.85	1.00	6866.61	4.23	0.0398

$rt \sim Polarity * Relation + (1|Item) + (1|Subject)$

Table 5 Experiment 1 - Maximal model reduced until convergence with cloze probability as covariate

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Polarity	85387.90	85387.90	1.00	2874.09	3.21	0.0735
Relation	35606.72	35606.72	1.00	75.64	1.34	0.2513
Cloze.Percentage	28678.23	28678.23	1.00	75.26	1.08	0.3028
Polarity:Relation	325.22	325.22	1.00	11242.33	0.01	0.9120
Polarity:Cloze.Percentage	139690.98	139690.98	1.00	11202.68	5.24	0.0221
Relation:Cloze.Percentage	188626.83	188626.83	1.00	76.06	7.08	0.0095
Polarity:Relation:Cloze.Percentage	22598.68	22598.68	1.00	11170.36	0.85	0.3571

$rt \sim Polarity * Relation * Cloze.Percentage + (1 + Relation|Item) + (1 + Polarity|Subject)$

Table 6 Experiment 2 - ANOVA table provided by the *lmerTest* package for the same analyses as in the main text

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Polarity	185606.04	185606.04	1.00	16295.49	4.28	0.0385
Relation	3787805.83	3787805.83	1.00	16296.77	87.37	0.0000
Polarity:Relation	167847.98	167847.98	1.00	16295.55	3.87	0.0491

$rt \sim Polarity * Relation + (1|Item) + (1|Subject)$

Table 7 Experiment 2 - Maximal model reduced until convergence

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Polarity	182975.71	182975.71	1.00	15734.07	4.28	0.0386
Relation	818165.24	818165.24	1.00	46.37	19.14	0.0001
Polarity:Relation	165304.56	165304.56	1.00	15731.84	3.87	0.0493

$rt \sim Polarity * Relation + (1 + Relation|Item) + (1 + Relation|Subject)$

Table 8 Experiment 2 - Maximal model reduced until convergence with cloze probability as covariate

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)
Polarity	9836.90	9836.90	1.00	15830.88	0.23	0.6314
Relation	11251.58	11251.58	1.00	44.65	0.26	0.6104
Cloze.Percentage	3012.13	3012.13	1.00	45.68	0.07	0.7918
Polarity:Relation	62.25	62.25	1.00	15930.69	0.00	0.9696
Polarity:Cloze.Percentage	914.51	914.51	1.00	15642.45	0.02	0.8837
Relation:Cloze.Percentage	170715.08	170715.08	1.00	44.50	3.99	0.0518
Polarity:Relation:Cloze.Percentage	17406.98	17406.98	1.00	15946.44	0.41	0.5234

$rt \sim Polarity * Relation * Cloze.Percentage + (1 + Relation|Item) + (1 + Relation|Subject)$

References

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The wacky wide web: A collection of very large linguistically processed web-crawled corpora. *Language Resources and Evaluation, 43*, 209–226.

Bates, D. (2005). Fitting linear mixed models in r. *R news, 5*, 27–30.

Beltrán, D., Liu, B., & de Vega, M. (2021). Inhibitory mechanisms in the processing of negations: A neural reuse hypothesis. *Journal of Psycholinguistic Research, 50*, 1243–1260.

Beltrán, D., Morera, Y., García-Marco, E., & Vega, M. d. (2019). Brain inhibitory mechanisms are involved in the processing of sentential negation, regardless of its content. evidence from eeg theta and beta rhythms. *Frontiers in psychology*, (p. 1782).

Braun, B., & Tagliapietra, L. (2010). The role of contrastive intonation contours in the retrieval of contextual alternatives. *Language and Cognitive Processes, 25*, 1024–1043.

Capuano, F., Dudschig, C., Günther, F., & Kaup, B. (2021). Semantic similarity of alternatives fostered by conversational negation. *Cognitive Science*, *45*, e13015.

Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. Harcourt Brace Jovanovich New York.

De Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior research methods, 47*, 1–12.

Dennison, H., & Schafer, A. J. (2017). Processing intonationally implicated contrast versus negation in american english. *Language and Speech, 60*, 174–199.

Givón, T. (1978). Negation in language: Pragmatics, function, ontology. In *Pragmatics* (pp. 69–112). Brill.

Gotzner, N., & Spalek, K. (2019). The life and times of focus alternatives: Tracing the activation of alternatives to a focused constituent in language comprehension. *Language and Linguistics Compass*, *13*, e12310.

Günther, F., Dudschig, C., & Kaup, B. (2015). Lsafun-an r package for computations based on latent semantic analysis. *Behavior research methods, 47*, 930–944.

Horn, L. (1989). *A Natural History of Negation*. The University of Chicago Press.

Husband, E. M., & Ferreira, F. (2016). The role of selection in the comprehension of focus alternatives. *Language, Cognition and Neuroscience, 31*, 217–235.

Kaup, B. (2009). How are pragmatic differences between positive and negative sentences captured in the processes and representations in language comprehension? *Semantics and Pragmatics: From Experiment to Theory*, (pp. 162–185).

Kaup, B., Ludtke, J., & Zwaan, R. A. (2005). Effects of negation, truth value, and delay on picture recognition after reading affirmative and negative sentences. In *Proceedings of the Annual Meeting of the Cognitive Science Society*. volume 27.

Kaup, B., Lüdtke, J., & Zwaan, R. A. (2006). Processing negated sentences with contradictory predicates: Is a door that is not open mentally closed? *Journal of Pragmatics, 38*, 1033–1050.

Kaup, B., Yaxley, R. H., Madden, C. J., Zwaan, R. A., & Lüdtke, J. (2007). Experiential simulations of negated text information. *Quarterly journal of experimental psychology, 60*, 976–990.

Kaup, B., & Zwaan, R. A. (2003). Effects of negation and situational presence on the accessibility of text information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 439.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods, 42*, 627–633.

Kruszewski, G., Paperno, D., Bernardi, R., & Baroni, M. (2016). There is no logical negation here, but there are alternatives: Modeling conversational negation with distributional semantics. *Computational Linguistics, 42*, 637–660.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmertest package: tests in linear mixed effects models. *Journal of statistical software, 82*, 1–26.

Liu, B., Wang, H., Beltrán, D., Gu, B., Liang, T., Wang, X., & de Vega, M. (2020). The generalizability of inhibition-related processes in the comprehension of linguistic negation. erp evidence from the mandarin language. *Language, Cognition and Neuroscience, 35*, 885–895.

MacDonald, M. C., & Just, M. A. (1989). Changes in activation levels with negation. *Journal of experimental psychology: Learning, memory, and cognition*, *15*, 633.

Mayo, R., Schul, Y., & Burnstein, E. (2004). "i am not guilty" vs "i am innocent": Successful negation may depend on the schema used for its encoding. *Journal of Experimental Social Psychology, 40*, 433–449.

Orenes, I., Beltrán, D., & Santamaria, C. (2014). How negation is understood: Evidence from the visual world paradigm. *Journal of Memory and Language, 74*, 36–45.

Orenes, I., Moreno-Ríos, S., & Espino, O. (2022). Representing negated statements: when false possibilities also play in the mind. *Journal of Cognitive Psychology, 34*, 1052–1062.

Palan, S., & Schitter, C. (2018). Prolific. ac-a subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27.

Repp, S., & Spalek, K. (2021). The role of alternatives in language. *Frontiers in Communication*, *6*, 111.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics, 1*, 75–116.

Schindele, R., Lüdtke, J., & Kaup, B. (2008). Comprehending negation: A study with adults diagnosed with high functioning autism or asperger's syndrome. *Intercultural Pragmatics, 5*, 421–444.

de Vega, M., Morera, Y., León, I., Beltrán, D., Casado, P., & Martín-Loeches, M. (2016). Sentential negation might share neurophysiological mechanisms with action inhibition. evidence from frontal theta rhythm. *Journal of Neuroscience, 36*, 6002–6010.

Viechtbauer, W. (2010). Conducting meta-analyses in r with the metafor package. *Journal of Statistical Software, 36*, 1–48.

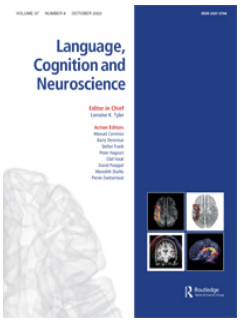
Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior, 4*, 7–11.

Yan, M., & Calhoun, S. (2019). Priming effects of focus in mandarin chinese. *Frontiers in psychology*, *10*, 1985.

Zimmermann, M. et al. (2008). Contrastive focus and emphasis. *Acta Linguistica Hungarica (Since 2017 Acta Linguistica Academica)*, *55*, 347–360.

Appendix C

Study 3: Capuano et al. (2022)



Language, Cognition and Neuroscience

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/plcp21

Using circles games to investigate the referential use of negation

Francesca Capuano, Carolin Dudschig & Barbara Kaup

To cite this article: Francesca Capuano, Carolin Dudschig & Barbara Kaup (2022): Using circles games to investigate the referential use of negation, Language, Cognition and Neuroscience, DOI: 10.1080/23273798.2022.2083645

To link to this article: https://doi.org/10.1080/23273798.2022.2083645

Published online: 07 Jun 2022.

LANGUAGE, COGNITION AND NEUROSCIENCE
https://doi.org/10.1080/23273798.2022.2083645

Routledge
Taylor & Francis Group

Using circles games to investigate the referential use of negation

Francesca Capuano, Carolin Dudschig and Barbara Kaup

Department of Psychology, University of Tübingen, Tübingen, Germany

ABSTRACT

Studies on the spontaneous production of negation suggest that it can be modulated by pragmatic principles of successful communication such as informativity and relevance. The present study investigates whether negation production is additionally modulated by a more general principle of effort minimisation. In a series of circles games, subjects were presented with pairs of circles and asked to complete a sentence that would allow a listener to identify one of the two circles. Negation was only produced when an affirmative description for the circle at issue was harder, i.e. there was no simple intuitive way to describe the circle's pattern. The length of the concurrent descriptions did not strictly influence the production of negation. The results suggest that the use of negation becomes more frequent as the effort to produce it decreases with respect to a concurrent affirmation, even at the cost of greater informativity of affirmation.

ARTICLE HISTORY
Received 2 August 2021
Accepted 18 May 2022

KEYWORDS
Negation; pragmatics; language production; economy

All natural languages allow for the construction of negative statements (Dahl, 1979; Horn, 1989). Despite its status being mostly marked (De Swart, 2009), negation is produced frequently in both written and spoken language. The apparent ease and naturalness with which speakers produce negation seemingly clashes with a long line of psychological research indicating that negative statements are a lot harder to process than affirmative statements (e.g. Carpenter & Just, 1975; Clark & Chase, 1972, 1974; Just & Carpenter, 1971, 1976; Wason, 1961; for an overview, see Kaup & Dudschig, 2020). The majority of these studies timed truth value judgments of negative vs. affirmative sentences with respect to world knowledge or a given visual world (e.g. *The star [is]/[is not] above the plus* coupled with a picture of a star standing above/below a plus sign in Clark Chase (1972)). The consistent finding across experiments was an increase in response times to negative sentences, that was unjustifiable on the sole basis of additional reading material (e.g. *not*).

Later on, considerable evidence was adduced to the idea that the difficulty of processing negation can be partly explained in terms of pragmatic felicity (Cornish, 1971; De Villiers & Tager Flusberg, 1975; Glenberg et al., 1999; Johnson-Laird & Tridgell, 1972; Lüdtke & Kaup, 2006; Nieuwland & Kuperberg, 2008; Wason, 1965). Specifically, the findings of these studies are all in line with the general idea that negation is particularly felicitous when it indicates deviations from expectations and/or corrects a false presupposition (Clark & Clark, 1977; Givón, 1978; Glenberg et al., 1999; Horn, 1989; Wason, 1965, 1972) that, if not necessarily held true by the listener, is at least relevant given the context (for a different view, see (Giora, 2006)). The processing asymmetry between affirmatives and negatives therefore seems to lie at the pragmatic level. According to Horn (1989), this asymmetry might originate from the interaction of two pragmatic principles: the Q principle (Quantity), which tends towards the maximal satisfaction of the listener's needs, and the principle of negative uninformativeness. Basically, as negation is usually less informative than affirmation, one must assume a context where it is just as informative as required (and no more than that, so as to not overload the listener with unsolicited information): those are precisely the contexts where the truth of the affirmative state of affairs is being considered.

In conclusion, negation comprehension seems to be sensitive to general pragmatic principles: listeners process negation faster if it is used in line with these principles. The more general dynamic is formalised in the Rational Speech Act model (Frank & Goodman, 2012; Goodman & Frank, 2016), a probabilistic model which at the basis of communication assumes a

CONTACT Francesca Capuano francesca.capuano@uni-tuebingen.de Department of Psychology, University of Tübingen, Schleichstraße 4, Tübingen 72076, Germany

recursive social reasoning between speaker and listener: the listener understands the utterance by reasoning on what the speaker would have said to fulfill their communicative goals while being rational (informative while also parsimonious). The model was shown to make successful quantitative predictions of pragmatic phenomena (Frank & Goodman, 2012). At least in the case of cooperative communication, it can be assumed that speakers do in fact behave rationally, and those contexts that are felicitous for negation comprehension should reflect negation usage when participants produce it in everyday life. Most work on negation though focused on the comprehension, as well as the completion of sentences that already encode polarity: there is very little evidence for when negation is spontaneously produced. In the following paragraphs we will briefly summarise previous studies that investigated negation production and their main findings.

As far as we know, Nordmeyer and Frank (2015) is the only study that established a direct link between negation production and comprehension, while explicitly relating the felicity of negation to general pragmatic principles. The authors showed that negation is spontaneously produced more often as it becomes more relevant and more informative. In each trial of their experiment, they presented subjects with a depiction of four characters. The trials differed in the proportion of characters holding the same, two recognisable objects (e.g. 0/4 or 1/4 or 2/4 or 3/4 or all characters were holding two apples). Apart from whether they were holding the objects, the characters presented within a trial were identical. Subjects were asked to complete sentences of the form *[NAME] has___* in reference to one of the four characters, which was highlighted by a red rectangle. The percentage of negative completions grew as the use of negation increased the probability of identifying the correct referent. In addition, the same depictions were paired with sentences of the form *[NAME] [has]/[has no] [ITEM]* and presented to a different batch of participants for a truth value judgment task to investigate comprehension ease of these items. Crucially, reaction times to true affirmative and negative sentences were predicted by their surprisal, calculated from the production probabilities derived from the speakers' task. The finding strongly supports a pragmatic view on the nature of the processing difficulties of negation. At the same time, informativity is highlighted as a drive for the production of negation. As noted by Nordmeyer and Frank (2015), these results support a Gricean perspective on negation processing (Grice, 1975): speakers follow – and listeners expect them to follow – the conversational maxims of informativity and relevance.

Older studies investigating the spontaneous production of negation can also be examined in this light. Watson (1979) prompted children to freely describe an entity so that it could be differentiated from a comparison entity. Negation was produced more often than affirmation to describe the referent (i.e. a white horse) when it lacked a property with respect to the contrast item (e.g. a horse with spots) but close to never when the contrast item differed along an attribute dimension (e.g. a black horse). This suggests that it can be used strategically when affirmation is less informative for the sake of the speech act.

Relatedly, Beltrán et al. (2008) show that negation is spontaneously produced when speakers might not have enough information to grant affirmation. Practically, negation is produced more often when an affirmative description of the state of affairs is not available. This availability was modulated by the use of bipolar vs. non-bipolar attributes (e.g. *big/small* vs. *red/green/blue/…*), as well as by whether the alternative state of affairs was explicitly mentioned in the context. In Experiment 1, participants were asked to complete short narratives describing the information reported on a source as erroneous (e.g. *In a magazine there was some wrong information. It talked about the [size/color] of a car. Juan realised that the information mistakenly stated that the car was [big/red]. In fact, the car___*). In Experiment 2, the alternative attribute was explicitly stated (e.g. *They discussed if the car was [big or small]/[red or green].*). More negative completions were produced in Experiment 1 in the non-bipolar condition, whereas the difference was absent in Experiment 2. Therefore, the effect seems attributable to the availability of the state of affairs rather than the attribute type. Producing a negation avoids the violation of the maxim of quality: the subjects lack adequate evidence for the actual state of affairs.

Taken together, the reported studies suggest that negation production seems to align to general pragmatic principles. The focus of these studies has been specifically on the role of classical Gricean principles underlying successful communication (such as informativity and relevance), which are primarily concerned with information content and the derivation of conversational implicatures. In the present study, we will focus instead on the role of a more general principle of effort economy underlying a wider range of human behaviours (Zipf, 1949), understood -- in the case of linguistic communication -- as the preservation of the cognitive effort of the speaker to retrieve and produce an utterance (for a discussion of economy as a broader principle, see (Carston, 2005)). We examine whether negation production can be modulated by speaker

economy, when the more economical form is not decidedly expected to be the affirmative option. In other words, speakers might be induced to produce negation not only on the basis of informativity and relevance expectations, but also following a general principle of least effort. We asked whether negation production can be modulated alone by the ease of production with respect to an alternative affirmation that serves the same communicative purpose. Subjects were presented with pairs of circles filled in varying patterns. They were asked to complete descriptions that would identify one of the two circles: asserting a property of the target circle would be as informative as negating the property of the concurrent circle for the sake of identifying the correct one. In each trial, the patterns of the two circles differed on the complexity of their affirmative description. The expectation is that, as long as it can fulfill the purpose of the speech act (identifying the correct circle), negation can be produced strategically based on economy considerations. Economy -- the preservation of cognitive effort to produce an utterance -- is initially operationalised on a very intuitive level: we distinguish between utterances that we *feel* are hard to produce from those that are easy. A preliminary investigation is then carried out to explore what might specifically index this production effort.

1. Circles game 1

1.1. Participants

Data were collected from 50 participants on Prolific (www.prolific.co) (15 male, 34 female and 1 other; age (in years): *mean* = 30.08, *sd* = 7.31). The task was estimated to take 5 minutes and was rewarded with 0.57£. The native language of the participants was set to English. At the beginning of the task they gave informed consent.

1.2. Materials

The items were 18 pairs of circles, each filled in a different pattern. For each pair, one circle was filled in a pattern that is easily nameable, the other with a pattern estimated hard to name. In the first three experiments, the material construction followed the principle that, for the *hard* to name condition, there was no intuitive single word to name that pattern. The circles were all displayed in the same size. An arrow would point at one of the two circles. Depending on the circle the arrow was pointed to, the item would vary on Target Difficulty (*easy* vs. *hard* to name). For an example stimulus, see Figure 1.

1.3. Design and procedure

After accepting the consent form, subjects were enquired on their age, gender and native language. They were then presented with the following instructions:

In this study you will see pairs of circles on the screen. Imagine you are talking to someone that can also see those circles and you want to draw their attention on the circle indicated by the arrow. Your task is always to continue the sentence 'Look at the circle…', such that you think that person would be able to identify the correct one. IMPORTANT: You CAN'T refer to the position of the circle (e.g. 'Look at the circle on the right/left'). Each time you will have 30 seconds to provide your answer.

After the instructions, the experimental trials started. Each participant was presented with all 18 items, each randomly assigned either to the *easy* or to the *hard* Target Difficulty condition. The arrow appeared simultaneously with the circles. Subjects had to enter their answer in a free text field before they could proceed to the next item. The experiment could be completed from pc/laptop with a standard keyboard. The order of presentation of the items, as well as the position of the circles relative to each other (left/right) in each pair were randomised.

1.4. Results

Participants who stated their native language was other than English were excluded from the analysis ($n = 1$).

Four raters (student assistants from our department) were shown the answers provided by the subjects, together with the corresponding stimuli deprived of the arrow. They evaluated the answers on:

- *Referent Position*: raters were asked *Which circle do you think the participant is referring to? [Right/Left/No idea]*. Only those datapoints were retained where at least three out of four raters agreed on either *Left* or *Right*, and the rated position actually corresponded to the position indicated by the arrow.
- *Grammaticality*: raters were asked whether the resulting sentences were grammatical. Answers rated as ungrammatical by at least two raters were excluded from the analysis.
- *Position Mentioned*: raters were asked *Does the sentence refer to the position of the circle? [yes/no]*. We kept only those answers for which at least three raters agreed on *no*.

Subjects left with less than 9 datapoints were excluded from the analysis. The answers were then

Figure 1. Example of an item in the *hard* to name Target Difficulty condition. On the left, the *easy* to name circle. On the right, the *hard* to name circle.

sorted by Polarity (whether they employed negation or not).

The cleaning procedure left us with 33 subjects. This drastic reduction in the number of subjects and therefore the amount of datapoints after the cleaning procedure is mostly attributable to the decision to retain only grammatical answers. Ungrammatical answers did not just simply contain typos but were often indicative of the subjects' non-compliance with the instruction to complete the given sentence (e.g. *Look at the circle …i like the wavey lines*). As a sanity check, we manually inspected all completions of Circles Game 1 that were excluded through the grammaticality check. Out of 314 completion, 114 were identified as particularly problmatic as they were constituted only by a noun phrase (e.g. *waves*). Even when additionally excluding those, the results did not change. As a further sanity check, the analyses to test our main hypothesis across all experiments presented in this paper were rerun on the data obtained by skipping the grammaticality check, which equally did not affect the pattern of the results and will therefore not be discussed any further.

13 out of 33 participants used negation. None of these used negation in the *easy* condition. The absence of observations in the "*easy* -" condition did not allow us to reliably fit a Generalised Linear Mixed-Effects Model to the data. Therefore, we run a Chi-Square test of independence to examine the relationship between Polarity and Target Difficulty. The relation between these variables was significant ($\chi^2(1) = 61.69$, $p < .001$). Negation was more likely to be produced in the case of *hard* items. Table 1 reports the overall counts per condition. Figure 2 plots the frequency of subjects by proportion of negative completions provided.

As hypothesised, negation can be produced in contexts where it becomes more economical relative to a competing affirmation. In fact, negative completions were only provided when the referent was *hard* to describe in affirmative terms.

2. Exploratory analysis -- part I

In this study, economy was operationalised as the ease to retrieve a sentence completion. One can easily consent that circles in the *easy* condition are less

Table 1. Circles game 1. Contingency table.

	–	+	Total
Easy	0	254	254
Hard	48	164	212
Total	48	418	466

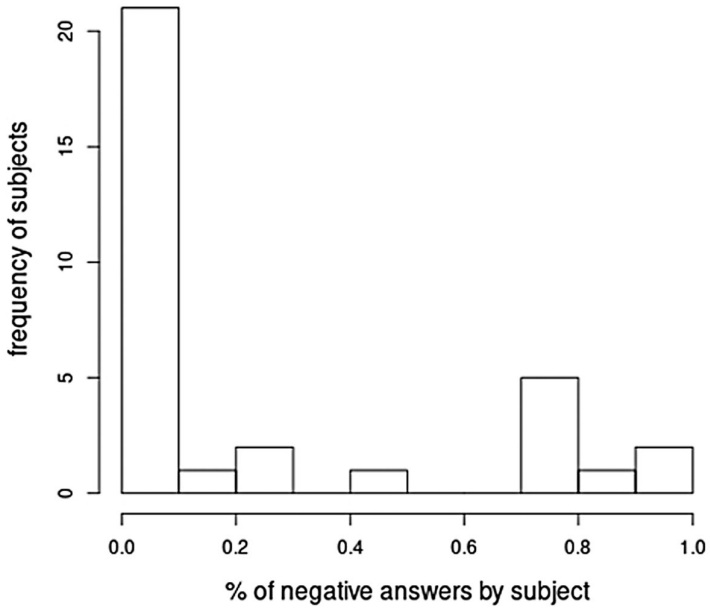


Figure 2. Circles game 1. Percentage of negative answers in the hard condition by subject.

effortful to describe than the ones in the *hard* condition, but the manipulation is only defined at a very intuitive level. As a first step to more clearly define what makes a sentence more cognitively effortful to produce (i.e. more or less economical), we explored different potential indices of production effort and -- whenever possible -- how they align with our main results.

We had 20 subjects (10 male and 10 female; age: *mean* = 33.45, *sd* = 10.44) describe each individual circle that was used in the main studies by completing the sentence *Look at the circle…* in 30 seconds. Next, they were asked to rate how hard it was to describe the circle, on a scale from 1 (= Very Easy) to 7 (= Very Hard). The data was collected on Prolific. Four dependent measures were considered: 1) perceived difficulty (the difficulty rating on the Likert scale); 2) writing time (*ms* from the beginning of the trial to the click to proceed to the next one); 3) completion length (number of characters employed); 4) average log frequency of the content words of the completion. To calculate 4), raw frequencies of the content words of each completion were extracted from the English Web Corpus *enTenTen* (Jakubíček et al., 2013) through SketchEngine (http://www.sketchengine.eu).

We tested for the main effect of Item Difficulty on each of the four variables (X) with linear mixed-effects models by comparing Model 1

$$X \sim ItemDifficulty + (1|Item) + (1|Subject) \quad (1)$$

with the baseline Model 2 through a likelihood ratio test.

$$X \sim +(1|Item) + (1|Subject) \quad (2)$$

First of all, *easy* circles were rated as actually easier to describe ($\beta = 3.01$, $\chi^2(1) = 75.12$, $p<.001$), confirming at the very least that the subjects shared our intuitions, but also that our manipulation corresponded to different levels of perceived difficulty after having actually produced these descriptions.

As for the other variables, 2) *easy* circles took less time to describe ($\beta = 5120$, $\chi^2(1) = 56.53$, $p<.001$); 3) *easy* circles were described with less characters ($\beta = 6.72$, $\chi^2(1) = 22.48$, $p<.001$); 4) descriptions of *easy* and *hard* circles employed on average equally frequent words ($\chi^2(1) < .001$, $p = .99$). Therefore, overall, perceived effort (effectively what we called *economy*) seems more correlated to the amount of time and characters employed to write, rather than with the frequency of the words employed.

We will get back to the exploratory analysis later in the text.

3. Circles game 2

In this second experiment, we aimed at replicating the findings of Circles Game 1, while increasing the availability of the alternative circle. The simultaneous presentation of the arrow and the circles in Circles Game 1 might have led subjects to more exclusively focus their attention on the selected circle from the very start. By

delaying the presentation of the arrow, it was hypothesised that subjects would be more likely to wholly perceive the visual context and formulate a description for the *easy* circle in advance, which would lead to an even higher rate of negation production: in that case, subjects could more easily reject the *easy* feature when the *hard* circle turned out to be selected.

3.1. Participants

Data were collected from 49 participants on Prolific (18 male and 31 female; age (in years): *mean* = 32.71, *sd* = 7.84).[1] Subjects who had already participated in the first task were excluded from further participation.

3.2. Materials

Same as in Circles Game 1.

3.3. Design and procedure

Same as in Circles Game 1, but the arrow appeared 5000 ms after the circles.

3.4. Results

The cleaning procedure left us with 34 subjects. Seventeen out of 34 participants used negation. Negation was more likely to be produced when referring to *hard* items ($\chi^2(1) = 71.34$, $p < .001$). The overall results are visualised in Table 2 and Figure 3.

The results of Circles Game 1 are replicated. Once more, negation is only produced in the *hard* condition, confirming that subjects only resorted to negation when its production disadvantage decreases compared to that of affirmation.

The proportion of negative completions produced did not appear to differ between Circles Game 1 and Circles Game 2. To test this, a Chi-Square test was performed on the *hard* data of Circles Games 1 and 2 to examine the relation between Polarity and Experiment ($\chi^2(1) = 0.03$, $p = .86$). 5000ms is a long delay, and we hypothesised that it might have allowed subjects to more likely formulate an affirmative description for both the *easy* and the *hard* circles. In the following experiment we shortened the delay in the presentation of the arrow to 500 ms, conjecturing that it might mostly increase the chances to formulate the *easy* description.

Table 2. Circles game 2. Contingency table.

	–	+	Total
easy	0	277	277
hard	54	173	227
Total	54	450	504

4. Circles game 3

4.1. Participants

Data were collected from 44 participants on Prolific (23 male and 21 female; age (in years): *mean* = 29.68, *sd* = 6.36). Subjects who had already participated in one of the previous tasks were excluded from further participation.

4.2. Materials

Same as in Circles Game 1.

4.3. Design and procedure

Same as in Circles Game 1, but the arrow appeared 500ms after the circles.

4.4. Results

The cleaning procedure left us with 25 subjects. 11 out of 25 participants used negation.

The relation between Polarity and Target Difficulty was again significant ($\chi^2(1) = 31.4$, $p < .001$). The overall results are visualised in Table 3 and Figure 4.

Crucially, the results of Circles Games 1 and 2 are replicated, as negation was only produced in the *hard* condition. A Chi-Square test was performed on the *hard* data of Circles Game 1 and 3 to examine the relation between Polarity and Experiment ($\chi^2(1) = 1.89$, $p = .17$). Like in Circles Game 2, the proportion of negative completions did not differ by experiment. Overall, the delay in the presentation of the arrow did not seem to affect the results. Possibly, the simultaneous presentation of the arrow and the circles in Circles Game 1 did not significantly drive the subjects' attention away from the full context. More factors might be at play and interact, as a longer SOA for the arrow might concurrently increase the chances to formulate an affirmative description for the *hard* item and the chances to have an *easy* description available to negate. The question needs further investigation.

In the next experiment, we examine more closely the concept of economy that might drive the production of negation. In Experiment 4 we manipulate economy more specifically in terms of the length of the expression needed to identify one of the two referents.

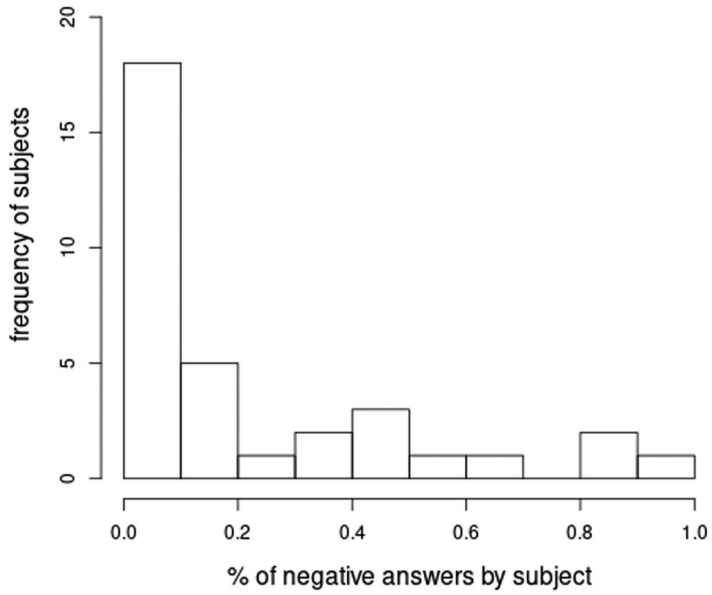


Figure 3. Circles game 2. Percentage of negative answers in the hard condition by subject.

5. Circles game 4

5.1. Method

Data were collected from 47 participants on Amazon Mechanical Turk (https://www.mturk.com/) (26 male, 20 female and 1 other; age (in years): *mean* = 36.4, *sd* = 11.04). The task was estimated to take 5–10 minutes and was rewarded with 1$. The location of the participants was set to the USA. The survey link instructions requested the participants to be native English speakers. At the beginning of the task they gave informed consent and were further enquired on their native language. Subjects who had already participated in one of the previous tasks were excluded from further participation.

5.2. Materials

The same 18 items from Circles Games 1–3 were employed and denoted as *pattern* items. Additionally, another 18 *length* items were created. These were also pairs of circles filled in a pattern, but the patterns were all repetitions of concrete entities. In the case of the *length* items, the naming difficulty of the two circles specifically differed in the length of the term needed to denote the entities in the circle: one was filled with a pattern of an entity with a long name, one with a pattern of an entity with a short name. See Figure 5 for an example of *length* item. Referents were chosen such that they had high depictability, and such that a corresponding depiction could be found, that was associated with the referent as univocably as possible (e.g. we avoided words such as *skyscraper*, whose depictions could easily be identified simply by *building*).

Table 3. Circles game 3. Contingency table.

	–	+	Total
Easy	0	186	186
Hard	30	151	181
Total	30	337	367

5.3. Design and procedure

Same as in Circles Game 1.

5.4. Results

The cleaning procedure left us with 18 subjects. The results are visualised in Table 4 and Figure 6.

5 out of 18 participants used negation. Surprisingly, one observation was produced in the *easy* - condition for one of the *length* items. We replicated the results of Circles Games 1–3 relative to the *pattern* items, in that subjects produced negative completions only when referring to the *hard* circles ($\chi^2(1) = 12.87$, $p < .001$). For the *length* items the Polarity of the answers did not differ by Target Difficulty (Fisher's exact test: $p = 1$).

The proportion of negative completions to the *pattern* items appears to be significantly higher than that to *length* items. To test whether the proportion of negative completions to *hard* items differed by Item

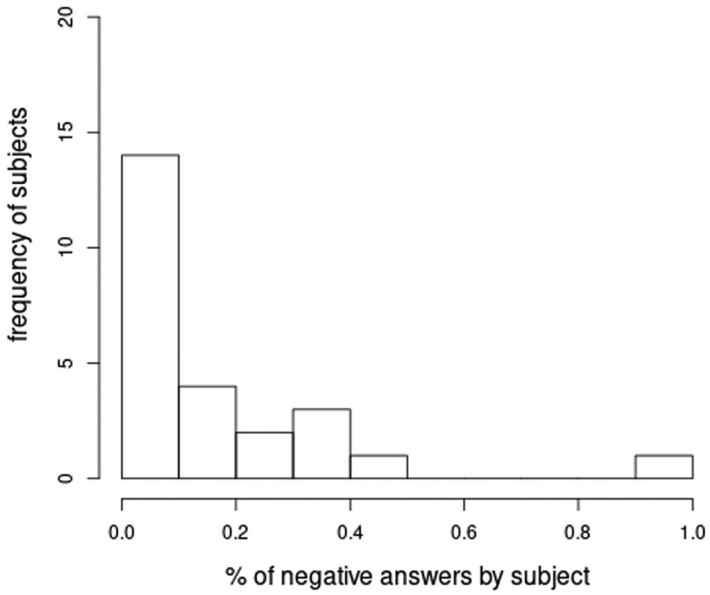


Figure 4. Circles game 3. Percentage of negative answers in the hard condition by subject.

Type (*pattern* vs. *length*), we performed a Chi-Square test ($\chi^2(1) = 9.4$, $p < .01$).

We conclude that negation can be produced on the basis of economy considerations, as the results of Circles Games 1 to 3 are replicated for the *pattern* items. As for the *length* items, two possibilities are worth considering. On the one hand, *hard* circles might not have been any harder to describe than *easy*

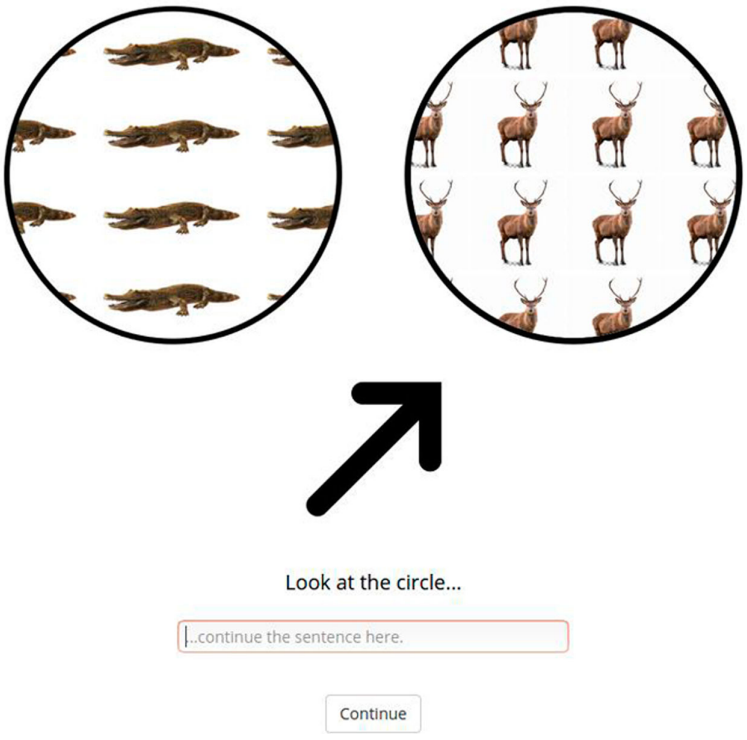


Figure 5. Example of a *length* item in the *easy* to name Target Difficulty condition. On the left, the *hard* to name circle (i.e. *crocodile*). On the right, the *easy* to name circle (i.e. *deer*).

Table 4. Circles game 4. Flattened contingency table.

		−	+
length	easy	1	142
	hard	2	142
pattern	easy	0	130
	hard	14	115

circles, therefore word/sentence length might not strictly quantify this production effort. A more comprehensive measure might include considerations on e.g. the ease of retrieval of words and/or syntactic complexity (see the exploratory analyses for a preliminary exploration of different possibilities). On the other hand, *hard* circles might have been harder than *easy* circles, but not enough to override other factors affecting the preference for affirmatives (see discussion).

6. Exploratory analysis -- part II

The same 20 subjects who completed the affirmative descriptions and the ratings from the first part of the Exploratory Analysis, analogously provided completions and ratings for the *length* circles employed in Circles Game 4. In the case of the *length* circles, *easy* circles were rated as equally easy to describe as *hard* circles ($\chi^2(1) = 0.18$, $p = .67$); 2) *easy* circles took less time to describe ($\beta = 1961$, $\chi^2(1) = 12.60$, $p<.001$) but the difference is smaller than in the case of the *pattern* items; 3) *easy* circles were described with less characters ($\beta = 4.89$, $\chi^2(1) = 15.6$, $p<.001$); 4) descriptions of *easy* circles employed more frequent words than descriptions of *hard* circles ($\beta = -0.99$, $\chi^2(1) = 9.60$, $p<.01$). A concise summary of the results of the exploratory descriptions of the individual circles, both *pattern* and *length*, can be found in Table 5.

Table 5. Exploratory analysis. Perceived difficulty, writing time, completion length and mean log frequency in the affirmative descriptions of the individual circles.

	Pattern Circles	Length Circles
Perceived Difficulty	easy < hard *** ($\beta = 3$)	easy = hard
Writing Time	easy < hard *** ($\beta = 5120$)	easy < hard *** ($\beta = 1961$)
Completion Length	easy < hard *** ($\beta = 6.72$)	easy < hard *** ($\beta = 4.89$)
Words Frequency	easy = hard	easy > hard ** ($\beta = -0.99$)

Differently from *pattern* items, in the case of *length* items, *easy* circles are not perceived to be any easier to describe than *hard* circles. This pattern of results matches the pattern of negation production across the main experiments: a difference in perceived difficulty within *pattern* circles corresponds to an increased production of negation, whereas the absence of such a difference in the *length* circles corresponds to a null effect. Therefore, *economy* as *perceived difficulty* seems to be a driving force behind negation production.

On the one hand, the *length* manipulation was effective, resulting in consistently shorter descriptions for *easy* items: these likely stem from the use of shorter nouns and, being accompanied by higher mean frequencies, might reflect the well-known

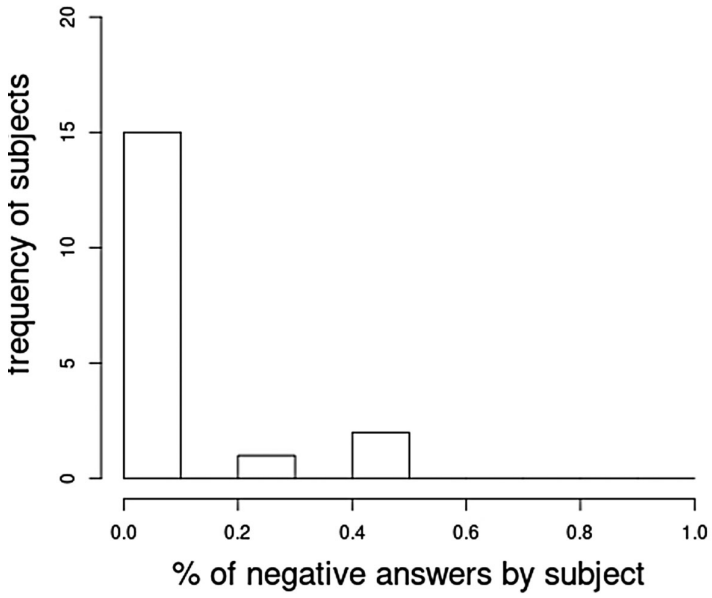


Figure 6. Circles game 4. Percentage of negative answers in the hard condition by subject.

Table 6. Exploratory analysis. Writing time by Polarity in the main experiments.

Circles Game 1	Circles Game 2	Circles Game 3	Circles Game 4
neg = aff	neg < aff *	neg = aff	neg < aff *

inverse correlation of word frequency and length (Zipf, 1949). On the other hand, this difference in length did not produce a difference in perceived difficulty. It did produce a difference in writing times, but this was greatly smaller than in the case of *pattern* items ($\beta = 1961$ vs. $\beta = 5120$) -- as confirmed by a significant interaction of Item Difficulty and Item Type ($\chi^2(1) = 19.87$, $p<.001$) -- and therefore likely a byproduct of increased word length. In fact, although differing in length and frequency, the *easy* and *hard* referents of the *length* circles are all very common and accessible concepts (e.g. *crocodile*, although a long word, is still very accessible). In the case of *pattern* items, greater difficulty might stem from the need to rephrase in common terms a concept for which no single word is immediately available. As a result of these observations, we can speculate that differences in *economy* are best reflected by differences in writing times.

Although this is by no means an inferential test but rather a coherence check, we observed that writing times *best* align with our -- and the subjects' -- intuitions on the different levels of production difficulty. This observation can -- cautiously -- feed the expectation that writing times should be somewhat more informative than completion length and average frequency as a measure of economy. We tested for differences in writing times between Polarity levels in the main experiments (Circles Game 1: $\beta = 1635$, $\chi^2(1) = 2.78$, $p = .10$; Circles Game 2: $\beta = 1934$, $\chi^2(1) = 5.05$, $p<.05$; Circles Game 3: $\beta = 2005$, $\chi^2(1) = 2.7$, $p = .10$; Circles Game 4: $\beta = 2859$, $\chi^2(1) = 5.73$, $p<.05$). Table 6 summarises these results.

In summary, affirmative completions always took the same or more time than negative completions. These exploratory results are compatible with the idea that, in those rare cases where subjects decided to employ negation, they did so because they had it available as a very advantageous option.

7. Discussion

Conversational negation displays pragmatic peculiarities and is not fully interchangeable with affirmation even when the amount of information conveyed is seemingly equated. Pragmatically felicitous contexts for negation have been identified, and include rejecting/correcting a false presupposition and referring to an exception. Nevertheless, studies on the spontaneous production of negation seem to show that it can follow general pragmatic principles such as informativity and relevance, without necessarily implying the rejection of false beliefs or reference to an exception.

The current study is a preliminary investigation into whether the spontaneous production of negation follows a principle of effort economy. We addressed whether negation can be spontaneously produced more often as it becomes increasingly economical compared to an affirmative expression fulfilling the same communicative purposes (i.e. when equating affirmation and negation on informativity). In the first three experiments economy was operationalised as the general ease to retrieve a description that would uniquely identify a referent. Negation was never produced when an affirmative description of the referent was *easy* to formulate, but it was produced strategically -- even though less often than affirmation -- when the affirmative description was *hard* and the same communicative purposes (i.e. identifying the referent within a visual context) could be fulfilled with a more economical negative expression. The result was replicated twice. Variations in the delay of presentation of the arrow that identified the referent did not produce any difference in the results. It was hypothesised that such a delay would increase the availability of a description for the context, with respect to which a negative description of the referent could be formulated, therefore leading to a higher proportion of negative productions. The hypothesis was not confirmed and further investigation is required to identify the underlying cognitive processes and their time course. In the fourth experiment economy was operationalised as the length of the description. The proportion of negative completions does not seem to differ by length of description, and it lies almost at 0 (1 negative observation in the *easy* and 2 in the *hard* condition).

We conclude that negation can and in fact is produced strategically on the basis of economy considerations. In fact, writing times align with the idea that, when they were produced, negatives were as economical as or more than the affirmative counterparts. If, as it is, affirmation was still largely preferred, this might be due to economy being overcome by different pragmatic considerations: identifying a referent on the basis of a feature it possesses still seems more pragmatically felicitous than identifying it on the basis of a feature it lacks. Although we assumed to have equated affirmatives and negatives on informativity, this might in fact not be the case. It has been argued that the informativity of a negative utterance is defined relative to a general discourse

QUD, where *polar* QUDs render negation particularly informative (Xiang et al., 2020): in Nordmeyer and Frank (2014), the context highlighted the absence/presence of a specific property (i.e. having/not having apples) as particularly salient, implicitly suggesting an underlying polar QUD (*Does Bob have apples?*). In this sense, the large prevalence of affirmations produced in our study despite both negation and affirmation being apparently equated on informativity might be due to the their deeper disparity in terms of underlying QUD (e.g. *How does the circle look like?* rather than *Does the circle have stripes?*) and therefore in terms of relevance. We can think of informativity as the update of prior beliefs: informative utterances reduce uncertainty in the state of affairs relative to a pertinent (i.e. *relevant*) question, with relevance being formalised by the concept of QUD. In our study, affirmatives and negatives were equated on informativity relative to the suggested communicative goal, in the sense that either would equally reduce the uncertainty in identifying the correct circle. Nevertheless, it seems like visual properties of the context might implicitly suggest other potential questions at issue.

In conclusion, the current study suggests that the effort of the speaker to retrieve and produce a negative utterance (i.e. strictly speaking *speaker economy*) contributes to its overall pragmatic felicity and modulates its production. Whereas the same is expected of any linguistic form, the production of negation is normally justified in terms of appropriateness of information content, seeing that, as a rule of thumb, negation is normally more effortful to produce than affirmation. Not only is speaker economy crucial, our results seem to suggest that greater pragmatic adequacy in terms of informativity is not a strictly necessary prerequisite for the production of negation: negation was spontaneously produced as it became a more and more economical option, even though affirmation was likely more informative for the question under discussion.

Note

1. Following a reviewer's suggestion, we later calculated power based on Circles Game 1, which resulted in only 60 observations (≈4 subjects) necessary to reach 80% power with $\alpha = .05$, suggesting we clearly have sufficient power to detect the interaction effect of Polarity and Target Difficulty ($w = .36$).

Data availability statement

The data that support the findings of this study are openly available in OSF at https://osf.io/jzeb7/?view_only=95bdd02ff280430d8a61af990c2ce413.

Disclosure statement

The authors report no conflict of interest.

Funding

This work was supported by the Collaborative Research Centre 833 (SFB833) *The Construction of Meaning*/Z2 project appointed to Barbara Kaup by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) [– SFB 833 – Project ID 75650358].

ORCID

Francesca Capuano http://orcid.org/0000-0003-1898-3255
Carolin Dudschig http://orcid.org/0000-0002-7041-7798
Barbara Kaup http://orcid.org/0000-0002-7400-0789

References

Beltrán, D., Orenes, I., & Santamaría, C. (2008). Context effects on the spontaneous production of negation. *Intercultural Pragmatics*, *5*(4), 409–419. https://doi.org/10.1515/iprg.2008.020

Carpenter, P. A., & Just, M. A. (1975). Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review*, *82*(1), 45–73.https://doi.org/10.1037/h0076248

Carston, R. (2005). Relevance Theory, Grice and the neo-Griceans: a response to Laurence Horn's 'Current issues in neo-Gricean pragmatics'. *Intercultural Pragmatics*, *2*(3), 303–319. https://doi.org/10.1515/iprg.2005.2.3.303.

Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology*, *3*(3), 472–517. https://doi.org/10.1016/0010-0285(72)90019-9

Clark, H. H., & Chase, W. G. (1974). Perceptual coding strategies in the formation and verification of descriptions. *Memory & Cognition*, *2*(1), 101–111. https://doi.org/10.3758/bf03197499

Clark, H. H., & Clark, E. V. (1977). *Psychology and language*. Harcourt Brace Jovanovich.

Cornish, E. R. (1971). Pragmatic aspects of negation in sentence evaluation and completion tasks. *British Journal of Psychology*, *62*(4), 505–511. https://doi.org/10.1111/j.2044-8295.1971.tb02065.x

Dahl, Ö. (1979). Typology of sentence negation. *Linguistics*, *17*(1–2), 79–106. https://doi.org/10.1515/ling.1979.17.1-2.79

De Swart, H. (2009). *Expression and interpretation of negation: An OT typology* (Vol. 77). Springer Science & Business Media. https://doi.org/10.9793/elsj.29.1_155

De Villiers, J. G., & Tager Flusberg, H. B. (1975). Some facts one simply cannot deny. *Journal of Child Language*, *2*(2), 279–286. https://doi.org/10.1017/s0305000900001100

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science (New York, N.Y.)*, *336*(6084), 998–998. https://doi.org/10.1126/science.1218633

Giora, R. (2006). Anything negatives can do affirmatives can do just as well, except for some metaphors. *Journal of Pragmatics*, *38*(7), 981–1014. https://doi.org/10.1016/j.pragma.2005.12.006

Givón, T. (1978). Negation in language: Pragmatics, function, ontology. In *Pragmatics* (pp. 69–112). Brill. https://doi.org/10.1163/9789004368873_005

Glenberg, A. M., Robertson, D. A., Jansen, J. L., & Johnson-Glenberg, M. C. (1999). Not propositions. *Cognitive Systems Research*, *1*(1), 19–33. https://doi.org/10.1016/s1389-0417(99)00004-2

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829. https://doi.org/10.1016/j.tics.2016.08.005

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill. https://doi.org/10.1163/9789004368811_003

Horn, L. (1989). *A natural history of negation*. The University of Chicago Press. https://doi.org/10.1016/0024-3841(90)90064-R

Jakubíček, M., Kilgarriff, A., Ková, V., Rychlý, P., & Suchomel, V. (2013). The tenten corpus family. In *7th international corpus linguistics conference CL* (pp. 125–127). Lancaster University.

Johnson-Laird, P. N., & Tridgell, J. M. (1972). When negation is easier than affirmation. *Quarterly Journal of Experimental Psychology*, *24*(1), 87–91. https://doi.org/10.1080/14640747208400271

Just, M. A., & Carpenter, P. A. (1971). Comprehension of negation with quantification. *Journal of Verbal Learning and Verbal Behavior*, *10*(3), 244–253. https://doi.org/10.1016/s0022-5371(71)80051-8

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Kaup, B., & Dudschig, C. (2020). Understanding negation: Issues in the processing of negation. In V. Déprez (Ed.), *The oxford handbook of negation* (Chapter 37, pp. 635–655). https://doi.org/10.1037/xge0000125

Lüdtke, J., & Kaup, B. (2006). Context effects when reading negative and affirmative sentences. In *Proceedings of the 28th annual conference of the cognitive science society* (Vol. 27, pp. 1735–1740).

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. https://doi.org/10.1111/j.1467-9280.2008.02226.x

Nordmeyer, A. E., & Frank, M. (2014). A pragmatic account of the processing of negative sentences. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 36, No. 36).

Nordmeyer, A. E., & Frank, M. C. (2015). *Negation is only hard to process when it is pragmatically infelicitous* [Unpublished manuscript]. Unpublished manuscript.

Wason, P. C. (1961). Response to affirmative and negative binary statements. *British Journal of Psychology*, *52*(2), 133–142. https://doi.org/10.1111/j.2044-8295.1961.tb00775.x

Wason, P. C. (1965). The contexts of plausible denial. *Journal of Verbal Learning and Verbal Behavior*, *4*(1), 7–11. https://doi.org/10.1016/s0022-5371(65)80060-3.

Wason, P. C. (1972). In real life negatives are false. *Logique et Analyse*, *15*(57/58), 17–38.

Watson, J. M. (1979). Referential description by children in negative form. *British Journal of Psychology*, *70*(2), 199–204. https://doi.org/10.1111/j.2044-8295.1979.tb01676.x

Xiang, M., Kramer, A., & Nordmeyer, A. E. (2020). An informativity-based account of negation complexity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(10), 1857–1867. https://doi.org/10.1037/xlm0000851.

Zipf, G. K. (1949). *Human behavior and the principle of least effort*. Addison-Wesley.