Connectedness: Papers by and for Sarah VanWagenen
UCLA Working Papers in Linguistics 18: 189–211
Carson T. Schütze & Linnaea Stockall (eds.), 2014

A subclinical study of the cognitive resources underlying scalar implicature: A focus on scalar adjectives

E. Matthew Husband

Recent investigations of different types of scalar implicature have reported variable strength in comprehenders' commitment to their implicated meanings, perhaps related to the different properties each scale type has. While such discoveries threaten a unified view of scalar implicature, this paper proposes that some of this variability may arise from the different extragrammatical cognitive resources these scales draw upon as a consequent of processing their different types of representation. Providing initial evidence that cognitive resources play a role in scalar implicature, this paper reports a subclinical study of scalar implicatures triggered by scalar adjectives, examining how the functionality of an individual's social cognition, semantic memory, and executive function affects the cancellation of implicatures compared to entailments. It reports specific effects of individuals' functionality in social cognition and semantic memory on their judgments of implicature and entailment cancellations. The functionality of individuals' semantic memory and executive function also affected judgment differences between antonymic scale cancellations. This evidence supports the possibility that cognitive resources influence scalar implicature, and that individual variability in these resources must be taken into account when considering the mechanisms of sentence processing.

Keywords autism quotient, cancellability, executive function, gradable adjectives, scalar implicature, semantic memory, subclinical

Introduction

Fluent conversation often requires comprehenders to draw appropriate inferences from a speaker's utterance in a regular and predictable manner, the rules of which are perhaps most significantly codified in Grice's (1975) maxims of conversation. Grice proposed that pragmatically competent speakers follow principles of rational cooperation, with the conversational maxims being particular instantiations of such principles, and that apparent deviation from maxims should act as cues that invite comprehenders to infer certain implicit meanings. The class of inferences arising from deviation of the maxim of quantity ("make your contribution as informative as is required"), known as scalar implicature, have been of particular interest to research in both linguistics and psycholinguistics.

Scalar implicatures can arise from a variety of different sources, including numerals, quantifiers, adjectives, rankings, and ad hoc categories, as evidenced in (1) respectively.

(1) a. Mary has two children.
b. Some snakes are poisonous.
c. This coffee is warm.
d. Juniors can live off campus.
e. John vacations in Greece.

The sentence in (1b), for instance, logically means that at least some (and possibly all) snakes are poisonous, but such a sentence is often taken to mean that some, but not all, snakes are poisonous. Under a Gricean account, this inference of an upper-bound, e.g. that not all snakes are poisonous, comes about because comprehenders assume that speakers are being cooperative and following the maxim of quantity. Having heard (1b), a comprehender knows that the speaker could have alternatively said that all snakes are poisonous. They also know that this alternative is logically stronger than what was actually said because it is true in fewer cases, and hence more informative. They then infer on the basis of the maxim of quantity that the speaker did not say this more informative alternative because the speaker could not truthfully say it; that is, the speaker was being as informative as he could be. Therefore the comprehender interprets the quantifier more restrictively to exclude the stronger meaning, leading them to take (1b) to mean that some but not all snakes are poisonous.[1]

The direction of these inferences are driven by the informativity, i.e. the logical strength, of other possible sentences the speaker could have said. Since its initial conception, the idea that some utterances are more informative than others has lead research to conclude that implicatures driven by the maxim of quantity are related to scales of informativity. Arguably the most explicit theory concerning the scales that underlie scalar implicature comes from Horn (1984). Horn proposed that informativity can be modeled in terms of entailments, in that stronger terms entail weaker ones and are therefore more informative, and that these entailment scales are stored as part of one's linguistic knowledge. Under this theory, scalar implicatures arise when a speaker uses a weaker alternative on a scale; that is, speakers choose to use weaker scalar alternatives when they are not in a position to truthfully say the stronger alternative, thus implicating that the stronger alternative is not true. This idea has since been extended to scales that arise from non-entailment sources (Hirschberg 1985), providing a unified account for the diverse types of cases that are all argued to be scalar implicatures. Returning to (1), for example, the following scales are thought to underlie the scalar implicature in each case: numerals <one, two, three>, quantifiers <some, many, all>, adjectives <warm, hot, scalding>, rankings <senior, junior, sophomore, freshman>, and ad hoc categories, <like Greek food, vacation in Greece, speak Greek>.

Whether the unified mechanism envisioned in Horn (1984) is a correct generalization, has however remained contentious. Chierchia (2004) and Levinson (2000), for instance, have proposed that context-independent scales, such as those arising from quantifiers and numerals, have a privileged status compared to context-dependent scales, such as those arising from rankings and ad hoc categories. Indeed, the scalar implicatures in (1d) and (1e)

[1] Additionally, comprehenders also judge whether a speaker knows the truth of the stronger alternatives, leading them to form an ignorance reading of the implicature in cases where they assume ignorance on the speaker's behalf (Van Rooij and Schulz 2004; Schulz and Van Rooij 2006). Although these factors may affect a comprehender's behavior (Goodman and Stuhlmüller 2013), I set them aside for the purposes of this study.

(that freshmen cannot live off campus, and that John does not speak Greek) are intuitively weaker and more fragile than those in (1b) and (1a). The experimental literature has also reported relative weakness in these cases (Doran et al. 2008; van Tiel et al. unpublished m.s.). Each of these studies, while employing different methods, finds variable strength in comprehender's commitment to implicated meanings. Taken together, these criticisms of a unified mechanism suggest either a diversity of mechanisms for scalar implicature or that other factors are at play in calculating scalar implicature.

One possible reason for the variation observed between different types of scalar implicatures may be the underlying cognitive resources that each type of scalar implicature draws upon during its computation. Although such extra-grammatical factors are a likely source of variation in the computation of scalar implicature, they have not been explored in much detail in the literature. One notable exception has been social cognition, a likely suspect given the central role social reasoning plays in the Gricean account (Bates 1976; Pijnacker et al. 2009; Surian 1996). A recent example of this comes from Nieuwland et al. (2010) which reported that a pragmatically related N400 response to an underinformative word was modulated by an individual's communicative score on the autism quotient scale. Assuming that this measure is related to some impairment of social cognition, it appears that calculation of scalar implicature relies on social reasoning skills which rapidly affect online processing.

Although central to the Gricean account, social cognition is not the only extra-grammatical faculty that is likely at play in the calculation of scalar implicature. Given the variety of scales that scalar implicature is thought to operate over, we might expect that different cognitive resources are required during implicature calculation depending on the type of scale being used. Different types of scales may, for instance, rely on the an individual's stored world knowledge and the relationships established between concepts, or they may require access to working memory, motivation, and goal-directed processes to be properly calculated – two cognitive resources discussed in broad terms as semantic memory and executive function, respectively.

Considering these two broad extra-grammatical resources, four theoretical possibilities could be considered. First, there may be implicatures that are so basic in their processing that they do not rely on the functionality of semantic memory or executive processes. Scalar implicatures of this sort would not draw upon world knowledge nor would they require comprehenders to remain engaged and motivated to draw proper inferences, perhaps because their scales are not based on conceptual relationships or are automatically triggered and do not overly burden memory resources.[2] Second, there may be implicatures that arise from representations stored in semantic memory but do not draw excessively from executive processes. Scalar relationships such as dimensional or prototypical concepts for instance could rely on the functionality of semantic memory to the extent they are stored in long-term memory and are accessed as part of basic lexical retrieval. Third, there may be implicatures that rely on the functionality of executive processes, but do not draw heavily on semantic memory. Implicatures with simple but unstored scalar alternatives could require high functioning executive processes if the computation of alternatives requires careful

[2]Speculatively, scalar implicatures of this type may fall within the kind that are context-independent (Chierchia 2004; Levinson 2000), perhaps, for instance, because they are completely internal to the language faculty or a primitive module that does not draw upon generalized external resources, although Marty et al. (2013) find that quantifier-triggered implicatures, a type of context-independent implicature, are disrupted by dual tasks that tap working memory.

coordination of attention and working memory. Finally, there may be implicatures that rely on the functionality of both semantic memory and executive processes. Calculation of implicatures based on these scales would likely draw both upon multiple stored semantic representations and require executive processes to coordinate these semantic representations within working memory.

This paper explores the second of these four possibilities — that there are scalar implicatures that rely on semantic memory, but not on executive function. Here, I focus on the scalar implicatures triggered by scalar adjectives. The scales underlying scalar adjectives likely depend on the structure of semantic memory because they are homogeneous stored semantic representations, but they also likely do not overly rely on high functioning executive processes because their homogeneous stored semantic representations do not require additional memory or attentional resources to coordinate their processing, prevent interference, etc. nor do they require further motivation and reasoning.

1 Scalar adjectives, scalar implicature, and cancellability

The study of scalar implicature has garnered particular interest in recent years from researchers in psycholinguistics, spawning a cottage industry of research examining the conditions which trigger scalar implicatures and the time course in which scalar implicatures are calculated (among other, Bott and Noveck 2004; Bott et al. 2012; Breheny et al. 2006; Hartshorne and Snedeker submitted; Huang and Snedeker 2009a; Grodner et al. 2010; Noveck and Posada 2003). Within this literature, an immediate methodological consideration is how to detect if and when comprehenders have calculated a scalar implicature given their implicit and inaudible nature. Many paradigms have relied on semantic expectancy/incongruency to detect processing difficulty resulting from the presence of a scalar implicature. Studying quantifier-triggered scalar implicatures, Bott and Noveck (2004), for instance, examined the reading times of underinformative sentences like (2a) to those of semantically incongruent sentences like (2b), finding that underinformative sentences too longer to accept than it took to reject semantically incongruent sentences.

(2) a. Some elephants are mammals.
b. Some elephants are insects.

Beltrama and Xiang (2012), studying scalar adjective-triggered implicatures, also used semantic incongruency in a self-paced reading task with a plausibility judgment. They manipulated whether the adjective was weak (*decent*), middle (*good*), or strong (*excellent*) and explored reading times on a critical word (*Harvard*, along with its spill-over) that would be incongruent given an inference for an upper-bound as in (3).[3] After reading the sentences, participants were asked if the final sentence was congruent with the previous sentence.

(3) Mark is a decent/good/excellent/bad student. That's why he has been accepted to Harvard for a Ph.D.

[3]The stimuli were in Italian, though an English gloss is given in (3). A further manipulation included a prior context sentence (*The competition for entering top programs is very tough.*), though this factor did not interact with participants' plausibility judgments or reading times.

Analysis of offline plausibility judgments found that only weak adjectives lowered the plausibility of the final sentence while middle and strong adjectives did not differ.[4] Beltrama and Xiang (2012) propose two possibilities for the lack of difference between middle and strong adjectives. First, special properties of scalar adjectives such as the lack of clear boundaries and the lack of upper bounds, especially for the middle adjective, may make the stronger alternatives of scalar adjectives relatively inaccessible. Second, strong adjectives may have a distinct semantic representation, such that strong adjectives are not actually scale-mates of the weak and middle adjectives (Morzycki 2012).[5]

Other online studies of scalar implicatures have made use of properties that are more particular to implicatures themselves to detect differences in processing. Implicatures, for instance, are cancellable, a core property distinguish them as a class of meaning from other classes, in which a previously implicated meaning can be explicitly suspended through a phrase like *if not*, *in fact*, or *actually* (Levinson 1983). Bezuidenhout and Morris (2004) relied on cancellability in their studies of quantifier-triggered scalar implicatures in an eye movement study, using sentences like (4) where a canceling phrase was preceded by *some*, triggering an upper-bound *not-all* implicature, or *the*, which has no such implication. The logic of the study was that if the quantifier triggered a scalar implicature, later cancellation of that implicature would take time as comprehenders revised their initial interpretative commitments. They found that participants gazed longer at a canceling phrase preceded by *some* compared to a canceling phrase preceded by *the*, suggesting that comprehenders rapidly generated a scalar implicature when reading *some*.

(4) Some/The books had color pictures. In fact all of them did, which is why the teachers liked them.

The study here also relied the cancellability of scalar implicatures. Interestingly in terms of scales, the literature has observed that cancellation of an implied meaning (one that is "up" the scale from the triggering word) is judged more acceptable than cancellation of an entailment (one that is "down" the scale from the triggering word). Also interesting are the differences in the strength of the acceptability judgments depending on the underlying scale. Quantifier scales like (5) yield strong and consistent differences, ranking scales like (7) yield weak and somewhat inconsistent differences, and adjectival scales like (6) fall somewhere in the middle.

(5) a. Some snakes, if not all of them, are poisonous.
b. ?* All snakes, if not some of them, are poisonous.

(6) a. The coffee is warm, if not hot.
b. ?? The coffee is hot, if not warm.

(7) a. Juniors, if not sophomores, can live off campus.

[4]Surprisingly, online measures failed to detect any processing differences given the adjective manipulation, suggesting that self-paced reading may not be sensitive enough to detect the indirect contextual incongruency of scalar implicature or that indirect contextual incongruency itself does not elicit a strong effect.

[5]The specific proposal here, that individual-variability in cognition resources affects scalar implicatures triggered by scalar adjectives, is not inconsistent with these proposals for scalar adjective-triggered implicatures. These proposals make particular representational claims about adjectives to account for the behavior of their scalar implicature, while the proposal here suggests a potential class of mechanisms, which likely interacts with representational nature of scales and influences the processing of scalar implicature.

b. ? Sophomores, if not juniors, can live off campus.

This study initiates a line of research investigating in part these differences in strength, beginning with scalar adjectives. Consider, for instance, the case of the temperature scale given in (8). A sentence like (9) which relies on this scale for its interpretation, entails *warm* (The lawyer's coffee is at least warm) and implicates *not scalding* (The lawyer's coffee is not scalding). Judgments of cancellation distinguish between these representations. Canceling the implicature in (9a) is acceptable, whereas canceling the entailment in (9b) is reported as odd. Finally, canceling to an antonym, which is not a scale-mate, as in (9c) is also judged to be odd.[6]

(8) <warm, hot, scalding>, cold

(9) The lawyer's coffee was hot.

a. The lawyer's coffee was hot, if not scalding and too bitter to drink.

b. ?? The lawyer's coffee was hot, if not warm and too bitter to drink.

c. ?? The lawyer's coffee was hot, if not cold and too bitter to drink.

The differences in judgments reflect a distinction in how scalar alternatives are represented in the interpretation of the basic clause containing the adjective *hot*. If extra-grammatical cognitive resources are required to encode these interpretative commitments, then individual-variability in the functionality of these resources should be reflected in the judgments elicited by different types of cancellation.

2 Cognitive resources and subclinical investigation

As discussed briefly above, different extra-grammatical cognitive resources likely underlie the processing of scalar implicatures, with different types of scales drawing on different cognitive resources. To investigate this possibility for scalar adjective-triggered implicatures, the study here employed a subclinical paradigm which correlates the behavior of healthy participants with clinical measurements of cognitive functions. Subclinical designs assume that the individual variability found within a healthy population by a clinical measurement, while not rising to a level of diagnosable disorder, is still correlated with functionality of the cognitive resource measured. Therefore, although participants in these studies are healthy and relatively unimpaired in their day to day cognitive functioning, the variation they display on clinical measures of cognitive resources, such as social cognition, semantic memory, and executive function each discussed below, allow us to detect whether these resources are drawn upon in performing a particular task.

[6] Beltrama and Xiang (2012) and van Tiel et al. (unpublished m.s.) both provide well reasoned reservations to explicitly providing scalar alternatives when studying scalar implicature, especially if one is concerned with the accessibility of implicit alternatives. The study here, however, is concerned with the cognitive resources used during the processing of scalar implicature, assuming that the alternatives are accessible and thus do generate scalar implicatures. Given this, I set these reservations aside, noting, however, that they should be kept in mind in interpreting any results.

2.1 Social cognition

Pragmatic reasoning, especially from the Gricean point of view, involves an understanding of the social rules that implicitly guide communication, be they due to reasoning about the general expectations of normal language use (e.g. Levinson 2000) or about comprehending the intentions of a speaker (e.g. Sperber and Wilson 1986). This type of reasoning is thought to be supported by social cognition, a cover term for the representations and processes individuals use to think about other individuals. Given its centrality to pragmatics, impairments to social cognition present a clear case in which pragmatic processing should be more difficult, for instance, because implicatures are not inferred when needed or because they are inferred when they should not be.

Delays and difficulties with social and communicative development are one type of social cognitive impairment, and are defining characteristics of autism. Individuals diagnosed with autism are known to exhibit problems with pragmatic reasoning, leading to inappropriate comments and difficulty comprehending non-literal language (Happé 1993; Wang et al. 2006) and indirect requests (Paul and Cohen 1985). Autism is also thought of as a spectrum disorder, with individuals displaying degrees of severity in autistic traits. The Autism-Spectrum Quotient (AQ) developed by Baron-Cohen et al. (2001) is used as a brief self-administered measurement of autistic traits in adults with normal intelligence. It examines five traits (social skill, attention switching, attention to detail, communication, and imagination) by asking individuals whether they agree or disagree with prompts like *I find it easy to "read between the lines" when someone is talking to me* and *I find it hard to make new friends*. Approximately half of the items are worded to elicit a disagree response from an individual diagnosed with autism while the other half of the items are worded to elicit an agree response.

The AQ has been used recently in psycholinguistic studies investigating scalar implicature (Nieuwland et al. 2010; Sikos et al. 2013), perspective taking (Grodner et al. 2012), and pragmatic reasoning (Xiang et al. 2013). The study by Xiang et al. (2013) used the AQ to detect the presence of pragmatic reasoning underlying illusory NPI licensing. They theorized that the high acceptability of illusory NPI licensing reported in the literature resulted from an over application of negative pragmatic inferences. They correlated individual AQ measures with reading times and acceptability judgments to licensed and unlicensed NPIs. They found that individuals with higher AQ measures correlated with lower acceptability of illusory NPI licensing, suggesting that they made fewer negative pragmatic inferences due to their lower pragmatic performance.

The study reported here follows Xiang et al. (2013) in its methodology in part. If scalar implicatures triggered by scalar adjectives rely on pragmatic reasoning skills, then we would expect AQ measures to correlate with the difference between cancellation of an implicature and cancellation of an entailment. The higher an individual's AQ score, the more similar they should find these types of cancellations since they are not generating the implicatures that would distinguish between "up" and "down" scale cancellations.

2.2 Semantic memory

In addition to relying on social cognition, scalar implicature also requires an understanding of alternatives and their relative strength. In some cases, like quantifier and numeral

triggered scalar implicature, the alternatives may be straightforward and tightly encoded, perhaps even internal to grammar itself (Chierchia (2004), though see (Chemla and Singh in press) for a discussion on the interplay of context in the grammatical approach to scalar implicature), while the alternatives to other types of scalar implicature may require a stronger appreciation of concepts and generic knowledge about the world. These other types of scalar implicature likely draw upon semantic memory during their processing.

At its core, semantic memory is thought of as memory for concepts and the relationship between them. Deficits to semantic memory are observed in patients with left hemisphere damage resulting in impaired semantic comprehension and is the core characteristic of semantic dementia. Patients diagnosed with semantic dementia have difficulty in word recall and have trouble pairing semantically associated images (Hodges et al. 1995). Semantic memory impairments are also thought to underlie thought disorder, described as a loosing of the associations between concepts, and are a characteristic of schizophrenia, Alzheimer's dementia, and psychosis (Doughty et al. 2009; Paula McKay et al. 1996).

The study here used two measures of semantic memory. The first is the Camel and Cactus Test (CCT), developed by Bozeat et al. (2000) as a more difficult version of the Pyramids and Palm Trees Test (Howard and Patterson 1992). This test assesses associative semantic memory using visual probes. Participants are given one target item (e.g. a camel) and are instructed to select one of four items that goes with the target (e.g. a cactus (correct), tree, sunflower, or rose). The second measure of semantic memory employed the Concrete and Abstract Word Synonym Test (CAWST), developed by Warrington et al. (1998). This test assesses semantic memory using linguistic probes. Participants are given one target word (e.g. *avarice*) and are instructed to choose the correct synonym for that word between two items (e.g. *greed* (correct) or *hunger*). The target words are either concrete or abstract, and come in graded difficulty, which allows for detection of minor degrees of impairment.

For scalar adjective-triggered implicatures, the distinction in cancellation of an "up" scale implicature from a "down" scale entailment rests on the encoding of scalar alternatives along the denoted dimension of the adjectives. This semantic knowledge may be a part of semantic memory itself, and if so we expect CCT and CAWST scores to correlate with the difference between cancellation of an implicature and cancellation of an entailment. Worse performance on semantic memory should lead to more similarity between these different types of cancellations as comprehenders begin to struggle with the semantic relationships between alternatives.

2.3 Executive function

In many cognitive tasks, multiple representations need to be coordinated, processed, and maintained to achieve particular behavioral goals. To achieve an intended interpretation and allow fluent conversation to continue to unfold, calculation of some types of scalar implicature may at times require this kind of control over cognitive processes, drawing upon an individual's executive function.

Executive function is a cover term for a variety of mechanisms that regulate and control cognitive processes, including working memory and attention, decision making, planning, monitoring, and problem solving. Deficits to executive function is observed in patients with frontal lobe damage, particularly in dorsolateral regions (Alvarez and Emory 2006;

Cummings 1993). Such impairments are also found in Alzheimers dementia, Parkinson's disease, and schizophrenia (Doughty et al. 2008; Dubois and Pillon 1996; Lafleche and Albert 1995).

As with semantic memory, this study used two measures for executive function. The Dysexecutive Questionnaire (DEX) was developed by Burgess et al. (1998) as part of a wider battery of tests examining executive function. Similar to the AQ, the DEX is a self-administered questionnaire. It examines three factors (behavioral, cognitive, and emotional (Wilson et al. 1998)) by asking individuals to rate how often a described behavior occurs on a five point scale from 'never' to 'very often'. Descriptions of behaviors include *I find it hard to stop repeating, saying, or doing things once started* and *I have difficulty thinking ahead or planning for the future*. The second measure of executive function used is Frontal Systems Behavior Scale (FrSBe) developed by Grace and Malloy (2001) (formally the Frontal Lobe Personality Scale (Grace et al. 1999)). This scale is also a self-administered behavior rating scale, assessing three factors (apathy, disinhibition, and executive dysfunction (Stout et al. 2003)) by asking individuals to rate how often a described behavior occurs on a five point scale from 'almost never' to 'almost always'. Descriptions of behaviors include *I am disorganized* and *I am able to plan ahead*.

Although executive function is likely to be involved in the calculation of any implicature, some types of scalar implicature may not require the careful coordination of multiple representations and processes that is likely to draw heavily on executive resources. Scalar implicatures triggered by scalar adjectives for instance may be of this type since only a single, centrally stored scale is called upon in their calculation. However, it is possible that representing the implicatures of scalar adjectives requires particular coordination of inhibition processes which suppress parts of the scalar dimension to generate the proper representation during implicature calculation. Therefore it is possible that scalar adjective-triggered implicatures require high functioning executive processes. If scalar adjective triggered scalar implicatures draw heavily on executive functions, we would expect DEX and FrSBe scores to correlate with the difference between cancellation of an implicature and cancellation of an entailment. Worse ratings of executive function should lead to more similarity between these types of cancellations as comprehenders encounter more interference from the scalar representation.

3 Experiment

3.1 Methods

Participants. 45 participants were recruited using Amazon's Mechanical Turk. Each participant was paid $4 for their participation.

Materials. We created 26 items consisting of three scalar alternatives and one antonym adjective. For each item we constructed 9 *if-not* cancellation sentences, pairing each of the adjectives with one another, as shown in Table 1. From these, 9 counterbalanced lists were created. 130 fillers (which included other types of corrections as well as unrelated sentence types) were also constructed and added to each list. The order of each list was pseudo-randomized, forming Section 1 of the stimuli.

Table 1: Example materials.

Scale	Cancellation		Sentence
Same	Implicature	1a	The lawyer's coffee was hot, if not scalding and too bitter to drink.
		b	The lawyer's coffee was warm, if not hot and too bitter to drink.
		c	The lawyer's coffee was warm, if not scalding and too bitter to drink.
Same	Entailment	2a	The lawyer's coffee was hot, if not warm and too bitter to drink.
		b	The lawyer's coffee was scalding, if not hot and too bitter to drink.
		c	The lawyer's coffee was scalding, if not warm and too bitter to drink.
Different	Antonym	3a	The lawyer's coffee was hot, if not cold and too bitter to drink.
		b	The lawyer's coffee was warm, if not cold and too bitter to drink.
		c	The lawyer's coffee was scalding, if not cold and too bitter to drink.

Section 2 consisted of 5 clinical questionnaires which were selected to detect subclinical variation in social cognition (AQ), semantic memory (CCT, CAWST), and executive function (DEX, FrSBe). To keep Section 2 short for participants, only a subset of items from each of these tests, chosen evenly from any subscales, was uses (AQ: 25 out of 50 items; CCT: 25 out of 64; CAWST: 26 out of 50 items; DEX: 20 out of 20 items; FrSBe: 24 out of 46 items). Scores for each test were calculated on the average, giving participant a score from 0 (high functioning) to 1 (low functioning) for the AQ, CCT, and CAWST, and a score from 1 (high functioning) to 5 (low functioning) for the DEX and FrSBe. These were added to the end of each list.

Procedure. In Section 1, participants were asked to rate the acceptability of sentences on a 1 (extremely unnatural) to 5 (extremely natural) scale. In Section 2, participants answered the clinical questionnaries.

Analysis. Acceptability ratings were modeled using mixed effects models with maximal random effects structure justified by the data (Baayen et al. 2008; Barr et al. 2013). Comparisons were Helmert coded for two contrasts: **Scale**, which compared cancellation to a scale mate (1a,b,c) and (2a,b,c) to cancellation to an antonym (3a,b,c), and **Cancellation**, which compared cancellation of an implicature (1a,b,c) to cancellation of an entailment (2a,b,c) within the same scale.[7] To test for correlations between clinical scores and acceptability, each clinical scale was added in as a fixed predictor in a separate model, with particular

[7]For clarity of presentation, the analyses in this paper focus on these two predictors, collapsing over a scalar distance condition. Analyses that include scalar distance do not change the overall pattern discussed below.

Table 2: Acceptability by sentence condition.

Condition		Average.	Stdev.
Cancellation	Implicature	3.82	1.08
	Entailment	2.97	1.20
	Antonym	1.74	1.01
Filler	Grammatical	4.47	0.79
	Ungrammatical	1.47	0.67

Table 3: Model estimates of acceptability for `acc` $\sim$ `condition + (1 | subject) + (1 + condition | item)` where condition is coded with Scale and Cancellation contrasts.

	Est.	Stdev.	t	p
Intercept	2.849	0.088	32.24	<.001***
Scale	1.103	0.064	17.33	<.001***
Cancellation	0.831	0.082	10.12	<.001***

* = $p < .05$, ** = $p < .01$, *** = $p < .001$

interest in the interaction coefficient between clinical scale and the Cancellation contrast.

3.2 Results

3.2.1 Acceptability

Average acceptability of each sentence condition is shown in Table 2. As expected, acceptability judgements recapitulated the traditional judgements reported in linguistic papers on implicature, with cancellation of an implicature rated more acceptable than cancellation of an entailment. Cancellation to an antonym scale showed lower acceptability. Interestingly, cancellation of an implicture was rated lower than typical grammatical filler sentences, suggesting that cancellation itself may lead to a minor drop in overall acceptability.

We first analyzed all the data together without any clinical score. The results found a significant effect of our manipulation, with both Scale and Cancellation having significant effect on sentence acceptability (t = 17.33 and t = 10.12 respectively). The results of this analysis can be found in Table 3.

3.2.2 Subclinical effects

Average scores on each subclinical scale and the correlation between each scale are given in Table 4. AQ scores were moderately correlated with both DEX and FrSBe scores. The CCT and CAWST scores were not significantly correlated with one anther; however, the DEX and FrSBe scores were themselves strongly correlated with one another. All other correlations were non-significant.

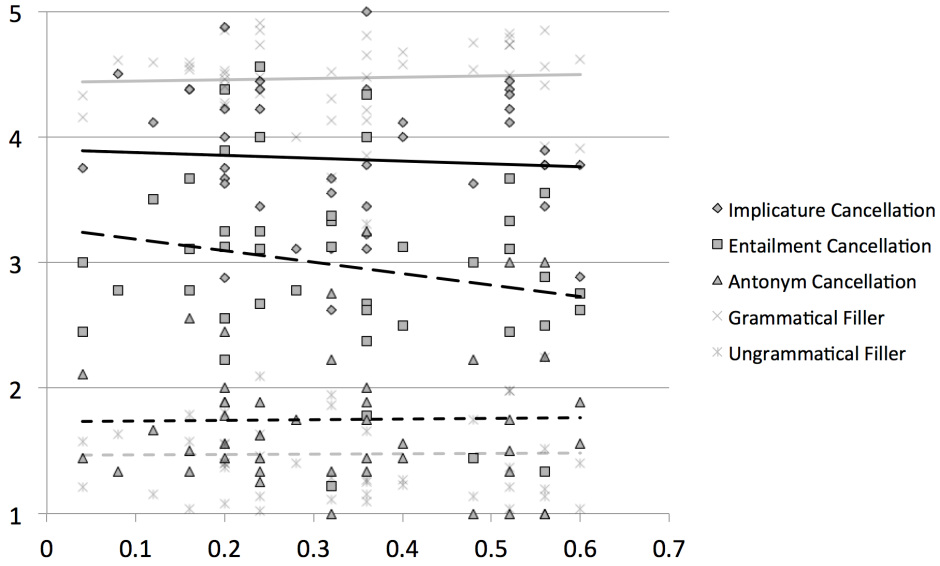
Each of the following sections examines the relationship between a cognitive resource and the acceptability of sentence conditions. The results of all subclinical analyses are given in Table 5.

Table 4: Average subclinical score and correlation between subclinical scales.

Clinical scale		Average	Stdev.	AQ	CCT	CAWST	DEX	FrSBe
Social	AQ	0.33	0.47	–	-.08	-.08	.41 **	.36 *
Semantic	CCT	0.23	0.42		–	.08	-.19	.04
	CAWST	0.77	0.42			–	-.13	-.28
Executive	DEX	2.06	0.97				–	.73 ***
	FrSBe	2.26	1.00					–

* = $p < .05$, ** = $p < .01$, *** = $p < .001$

Figure 1: Sentence acceptability by AQ score. Each data points represent a single subject's mean for that condition. A linear fit for implicature cancellation is given in a black solid line, for entailment cancellation in a black long dashed line, and for antonym cancellation in a black short dashed line. Grammatical and ungrammatical fillers are also presented, with linear fits in a light gray solid and short dashed line respectively.



Social cognition Subclinical impairment in social cognition as measured by AQ interacted positively with Cancellation ($t = 2.050$). As shown in Figure 1, this overall increase in the difference between cancellation of an implicature vs. an entailment was driven primarily by entailment cancellations. The more autistic traits a participant reported the more unacceptable they found cancellation of an entailment. There was no significant interaction between AQ score and Scale.

Semantic memory Subclincial impairment on semantic memory as measured by CCT did not significantly interact with Scale or Cancellation. Subclinical impairment on semantic memory as measured by CAWST interacted positively with Scale ($t = 2.698$) and Cancellation ($t = 2.017$). The significant Scale estimate, modeling an overall increase in the difference between cancellations on the same scale compared to the antonym scale, was driven primarily by the antonym cancellations. The lower a participant's semantic

memory measure was, the more unacceptable they judged the antonym cancellation. The significant Cancellation estimate, modeling an overall increase in the difference between cancellation of an implicature vs. an entailment, was driven by both sentence conditions. The lower a participant's semantic memory measure was, the more acceptable they judged the cancellation of an implicature and the more unacceptable they judged the cancellation of an entailment. Both of these effects can be seen in Figure 2B.

Executive function Subclinical impairment on executive function as measured by both DEX and FrSBe interacted negatively with Scale (DEX: t = -4.444; FrSBe: t = -3.924). As shown in Figure 3, this overall decrease in the difference between cancellations on the same scale compared to the antonym scale was driven by both sentence conditions. The higher a participant's DEX measure was, the less acceptable they judged cancellations of implicatures and the more acceptable they judged the cancellation to an antonym scale. The same was true for FrSBe measures, with the addition that the higher a participant's score was on the FrSBe, the less acceptable they also judged cancellations of entailments. There was no significant interaction between DEX or FrSBe and Cancellation.

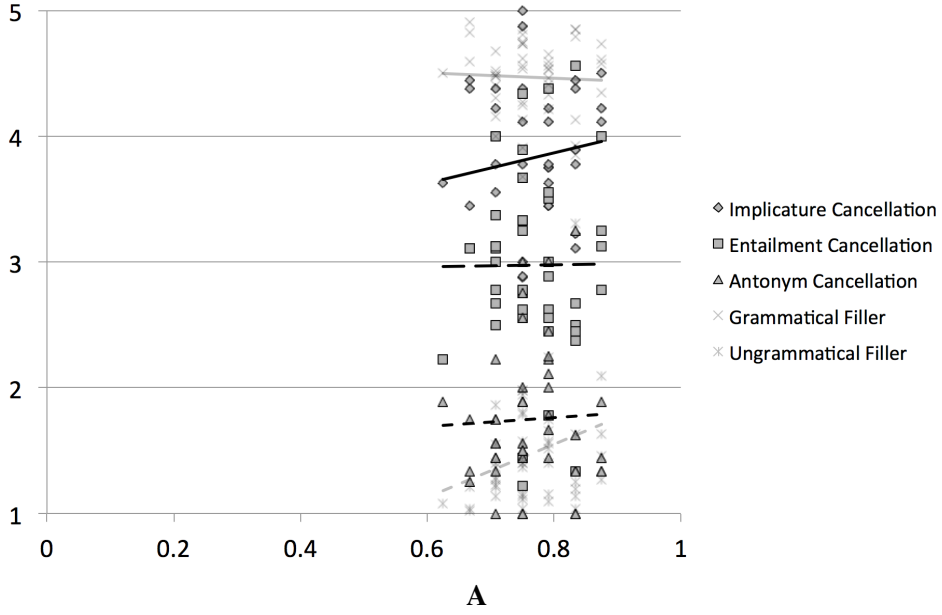
3.3 Discussion

Overall, this study finds evidence of extra-grammatical cognitive resources at work in the calculation of scalar implicatures triggered by scalar adjectives. Both social cognition and semantic memory resources differentially affected the acceptability of implicature and entailment cancellations, and semantic memory resources and executive function differentially affected the acceptability of cancellations between antonymic scales.

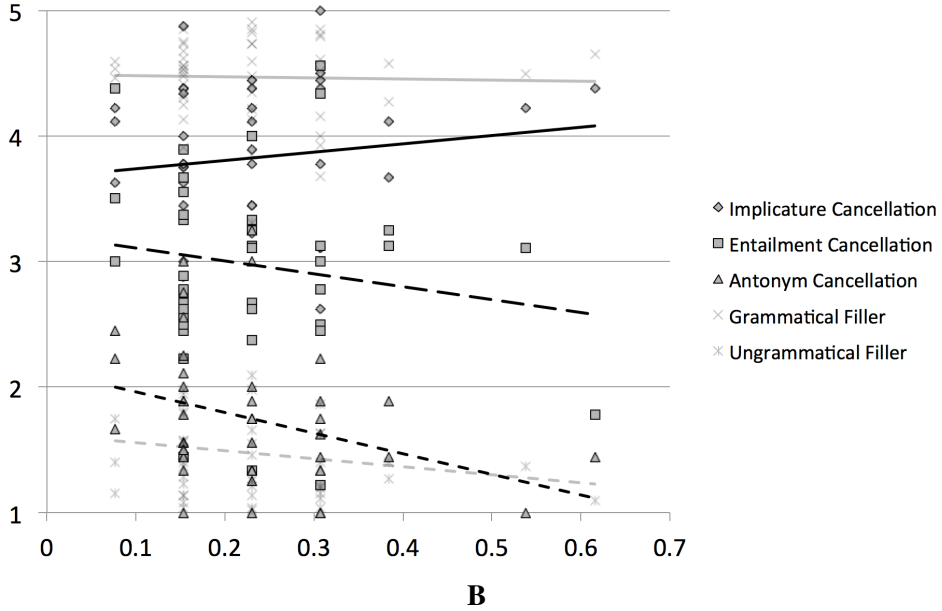
Interestingly, and counter to initial predictions, individuals with more impaired social cognition and semantic memory showed a greater, not lesser, distinction between implicature cancellations and entailment cancellations compared to their less impaired counterparts. While initially surprising, these results are sensible when one considers that acceptability judgments are a relatively late measure that taps into the final products of sentence comprehension. It could be that individuals with higher social cognition find entailment cancellation more acceptable because they more readily reason about why someone might have said such a sentence, and more often are able to construct a scenario in which such a sentence would be sensible and thus acceptable. For instance, an entailment cancellation like *The lawyer's coffee was hot, if not warm and too bitter to drink* might cue a comprehender to a more particularized implicature, say, that the speaker believes that lawyers are unable to be satisfied, as indicated by their flippant dislike of their beverage regardless of temperature. Individuals with lower social cognition, however, may take such sentences more on face value, regarding the oddness as mere oddness and not as an invitation to draw further inferences. They may, therefore, more quickly abandon such sentences without considering alternative interpretations that might have been intended by the speaker. Note, however, that only within-scale oddity appeared to invite rescuing inferences as social cognition did not have an effect on cancellations to antonym scales. It may be that these cross-scale sentences are considered too odd to rescue through social considerations even by high functioning individuals.

In a similar vein, individuals with high functioning semantic memory may have an easier

Figure 2: Sentence acceptability by CCT (A) and CAWST (B) scores. Each data points represent a single subject's mean for that condition. A linear fit for implicature cancellation is given in a black solid line, for entailment cancellation in a black long dashed line, and for antonym cancellation in a black short dashed line. Grammatical and ungrammatical fillers are also presented, with linear fits in a light gray solid and short dashed line respectively.

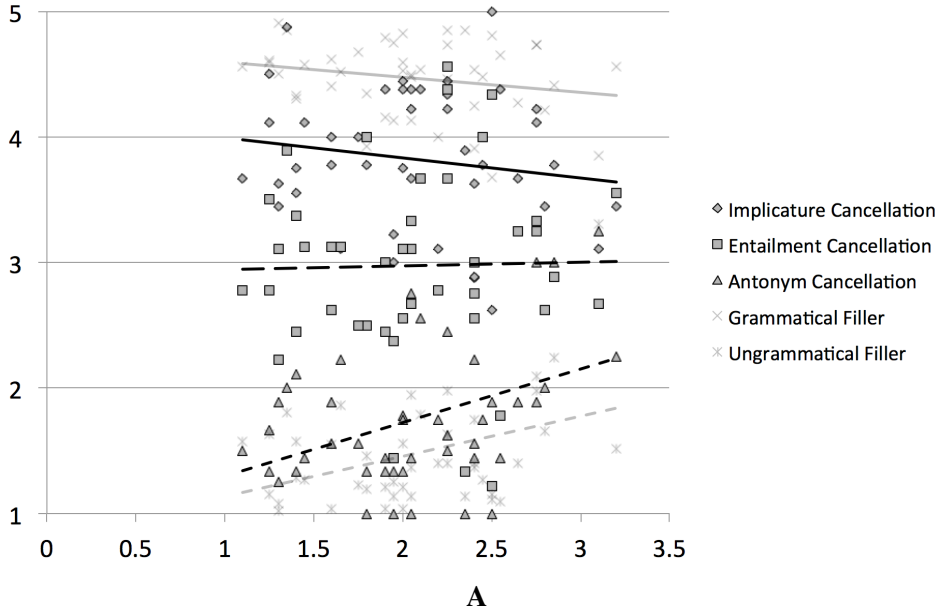


A

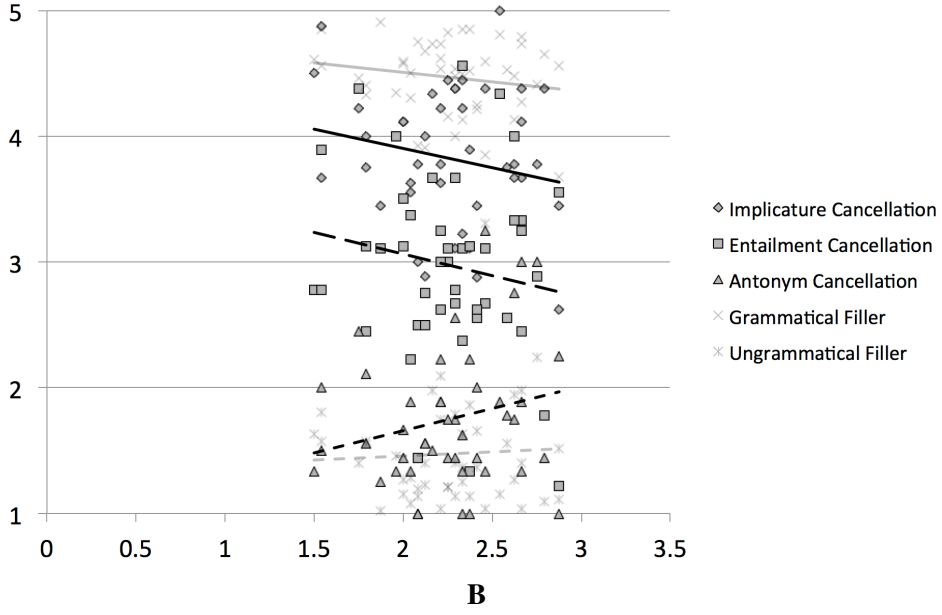


B

Figure 3: Sentence acceptability by DEX (A) and FrSBe (B) scores. Each data points represent a single subject's mean for that condition. A linear fit for implicature cancellation is given in a black solid line, for entailment cancellation in a black long dashed line, and for antonym cancellation in a black short dashed line. Grammatical and ungrammatical fillers are also presented, with linear fits in a light gray solid and short dashed line respectively.



A



B

Table 5: Model estimates of acceptability for `acc ~ subclinical*condition + (1 | subject) + (1 + condition | item)` where subclinical is the subclinical scale and condition is Helmert coded with Scale and Cancellation contrasts.

Subclinical Scale	Coefficient	Est.	Stdev.	t	p
AQ	Intercept	2.964	0.169	17.512	<.001***
	AQ	-0.345	0.435	-0.793	.428
	Scale	1.229	0.103	11.989	<.001***
	Cancellation	0.528	0.171	3.088	.002 **
	AQ:Scale	-0.383	0.256	-1.495	.135
	AQ:Cancellation	0.914	0.446	2.050	.040 *
CCT	Intercept	2.528	0.873	2.896	.004 **
	CCT	0.421	1.138	0.370	0.711
	Scale	1.079	0.508	2.125	0.034*
	Cancellation	0.044	0.899	0.049	.961
	CCT:Scale	0.032	0.660	0.049	.961
	CCT:Cancellation	1.032	1.173	0.880	.379
CAWST	Intercept	2.985	0.169	17.700	<.001***
	CAWST	-0.594	0.628	-0.946	.344
	Scale	0.876	0.107	8.164	<.001***
	Cancellation	0.533	0.173	3.080	.002 **
	CAWST:Scale	0.996	0.369	2.698	.007 **
	CAWST:Cancellation	1.306	0.647	2.017	.044 *
DEX	Intercept	2.665	0.287	9.273	<.001***
	DEX	0.090	0.132	0.685	.494
	Scale	1.786	0.178	10.047	<.001***
	Cancellation	1.050	0.330	3.180	<.001***
	DEX:Scale	-0.332	0.075	-4.444	<.001***
	DEX:Cancellation	-0.106	0.128	-0.828	.408
FrSBe	Intercept	3.183	0.438	7.260	<.001***
	FrSBe	-0.146	0.189	-0.776	.438
	Scale	2.074	0.255	8.119	<.001***
	Cancellation	0.698	0.446	1.565	.118
	FrSBe:Scale	-0.427	0.109	-3.924	<.001***
	FrSBe:Cancellation	0.057	0.192	0.299	.765

* = $p < .05$, ** = $p < .01$, *** = $p < .001$

time understanding the semantic relationships implicated by cancellation, perhaps because their semantic representations are richer. High functioning semantic memory may also give individuals the flexibility to understand more readily the relationship between *warm*, *hot*, and *scalding* as not just implicational but also degree-based. This may leave them more equipped to reconcile the oddness of an entailment cancellation by seeing these as relationships of degree than those whose semantic memory is more static and less specified. Semantic memory also affects the relationship to antonym scales. Since antonym relationships are encoded and retrieved from semantic memory, individuals with high functioning semantic memory again have a better ability to more carefully consider their relationship. In each case, this leads individuals with high functioning semantic memory to moderate their acceptability judgments, whereas individuals with low functioning semantic memory make more extreme judgments based on the initial naturalness of the condition.

These effects of semantic memory, however, were only found for the CAWST, a verbal synonym task. The CCT, which queries the semantic relationships between images, showed no significant effects (although visual inspection of Figure 2A appears to show a widening difference between implicature and entailment cancellations). There are at least two possibilities for the discrepancy between these two semantic memory measures. First, it could be that because the CCT taps semantic memory through a non-linguistic route, it is less sensitive to the semantic memory resources used in linguistic tasks. However, it could also be that unusually low CCT scores may have obscured any effects. In previous studies, the normal range for healthy controls is 56-65% correct while patients diagnosed with semantic dementia tend to fall below 48% (Adlam et al. 2010; Bozeat et al. 2000). The participants in this study, assumed to be healthy, fell within 12.5-37.5% correct, suggesting that the task itself was not well understood, perhaps because of the online presentation or unclear instructions. Follow up studies examining online processing will address this issue by bringing participants into the lab where a more standardized version of the CCT can be administered.

The effects of executive function present a different pattern from social cognition or semantic memory. Individuals with lower scores in DEX and FrSBe measures had lower acceptability scores for implicature and entailment cancellation and higher acceptability scores for antonym cancellation. In general, the lower an individuals score, the smaller the range of their acceptability scores, with judgments narrowing towards the middle of the acceptability scale. This suggesting that executive function may be related to making sharper distinctions between different representations, with high functioning individuals better able to detect the differences between the relative oddness of entailment cancellation and the outright contradiction of antonym cancellation. This could be related to an individual's motivation and deployment of attentional resources throughout the study, which would allow them to better monitor their responses on a trial to trial basis.

Conclusion

Taken together, these data argue that in addition to the role social cognition plays in the interpretation of scalar implicature, those scalar implicatures triggered by scalar adjectives also draw upon semantic memory resources, suggesting that semantically rich adjectival scales require high functioning semantic memory resources to properly act as a basis for

implicature calculation and reasoning about cancellation. Interestingly, executive function did not play a significant role in these particular types of implicatures, but instead affected the overall range of acceptability scores, suggesting that attention and working memory resources are not particularly called upon when the scales required are homogeneously stored as part of semantic memory.

Future work will extend this paradigm to other types of scales which are likely to reveal either relative insulation from or deep reliance on extra-grammatical cognition. On the one side, given previous research showing that exactly-interpretations of numerals do not show a processing cost (Huang and Snedeker 2009a; Marty et al. 2013) and develop earlier than other types of scalar implicatures (Papafragou and Musolino 2003; Musolino 2004; Huang and Snedeker 2009b), numeral scales may be a prime example of a scalar implicature insulated from the functionality of extra-grammatical semantic memory or executive function. An open question is whether quantifier scales are also of this variety, for while they are similar to numerals, they show processing costs and later development. On the other side, scale types like ranking and ad hoc scales are likely to trigger scalar implicatures that are highly reliant on the functionality of semantic memory and executive function given the abstract and indirect nature of the relationships that establish these scales.

Additional studies will also examine the real-time processing of these scalar implicatures. It seems very possible that the offline nature of acceptability judgments obscured and distorted the underlying effects of cognitive resources on implicatures and the comparison of their cancellation with the cancellation of entailments. I have proposed above that individuals with high functioning social cognition and semantic memory elaborate on cancellation, leading them to moderate their acceptability judgments of these sentences in comparison to those with low functioning social cognition and semantic memory. This suggests that early measures of processing may reveal a deeper initial similarity between individuals, with later processing uncovering the appreciation of the distinctions in meaning for high functioning individuals.

Further research would also be welcomed in a more precise understanding of what each of these clinical measurements are capturing in terms of cognition. Neither semantic memory nor executive function are single unified resources, and a better understanding of how impairments measured by these different clinical instruments match up with cognitive representations and mechanisms will improve our understanding both of their particular effects on language processing and also how studies like this can inform what the underlying representations and mechanisms actually at play are. This kind of joint venture will likely benefit both our understanding of cognition and of clinical disorder.

Regardless of the particulars, this study demonstrates that extra-grammatical cognitive resources must be taken into consideration when trying to understand the variability found within and across studies of scalar implicature. It is possible that the variation between different types of scalar implicature comes not from some diversity of mechanisms but rather from the variation within the cognitive systems that interact with a unified mechanism during processing. This research demonstrates the potential importance of cognitive resources to the computation of scalar implicature in particular, and to linguistic processing more generally.

Acknowledgements

I would like to thank Nick Angelides for help in preparing the stimuli for this study and St. Hugh's College for funding. This paper is dedicated to Sarah VanWagenen, whose sheer grit over these past few years serves as an example to us all. Sarah, you're a rockstar.

References

Adlam, Anna-Lynne R., Karalyn Patterson, Sasha Bozeat, and John R Hodges. 2010. The cambridge semantic memory test battery: Detection of semantic deficits in semantic dementia and alzheimer's disease. *Neurocase* 16:193–207.

Alvarez, Julie A., and Eugene Emory. 2006. Executive function and the frontal lobes: A meta-analytic review. *Neuropsychology Review* 16:17–42.

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

Baron-Cohen, Simon, Sally Wheelwright, Richard Skinner, Joanne Martin, and Emma Clubley. 2001. The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians. *Journal of Autism and Developmental Disorders* 31:5–17.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:255–278.

Bates, Elizabeth. 1976. *Language and context: The acquisition of pragmatics*. Academic Press New York.

Beltrama, A., and M. Xiang. 2012. Is good better than excellent? An experimental investigation on scalar implicatures and gradable adjectives. In *Proceedings of Sinn und Bedeutung*, ed. E. Chemla, V. Homer, and G. Winterstein, volume 17, 81–98.

Bezuidenhout, Anne L., and Robin K. Morris. 2004. Implicature, relevance and default pragmatic inference. In *Experimental pragmatics*, ed. D. Sperber and I. A. Noveck, 257–282. Palgrave Press.

Bott, Lewis, Todd M. Bailey, and Daniel Grodner. 2012. Distinguishing speed from accuracy in scalar implicatures. *Journal of Memory and Language* 66:123–142.

Bott, Lewis, and Ira A. Noveck. 2004. Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of Memory and Language* 51:437–457.

Bozeat, Sasha, Matthew A. Lambon Ralph, Karalyn Patterson, Peter Garrard, and John R Hodges. 2000. Non-verbal semantic impairment in semantic dementia. *Neuropsychologia* 38:1207–1215.

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

Burgess, Paul W., Nick Alderman, Jon Evans, Hazel Emslie, Barbara A Wilson, et al. 1998. The ecological validity of tests of executive function. *Journal of the International Neuropsychological Society* 4:547–558.

Chemla, Emmanuel, and Raj Singh. in press. Remarks on the experimental turn in the study of scalar implicature. *Language and Linguistics Compass* .

Chierchia, Gennaro. 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In *Structures and beyond*, ed. A. Belletti, 39–103. Oxford University Press Oxford.

Cummings, Jeffrey L. 1993. Frontal-subcortical circuits and human behavior. *Archives of Neurology* 50:873.

Doran, R., R. Baker, Y. McNabb, M. Larson, and G. Ward. 2008. On the non-unified nature of scalar implicature: An empirical investigation. *International Review of Pragmatics* 1:1–38.

Doughty, Olivia J., D. John Done, Verity A. Lawrence, Adil Al-Mousawi, and Kunle Ashaye. 2008. Semantic memory impairment in schizophrenia: Deficit in storage or access of knowledge? *Schizophrenia Research* 105:40–48.

Doughty, Olivia J., Verity A. Lawrence, Adil Al-Mousawi, Kunle Ashaye, and D. John Done. 2009. Overinclusive thought and loosening of associations are not unique to schizophrenia and are produced in Alzheimer's dementia. *Cognitive Neuropsychiatry* 14:149–164.

Dubois, Bruno, and Bernard Pillon. 1996. Cognitive deficits in Parkinson's disease. *Journal of Neurology* 244:2–8.

Goodman, Noah D., and Andreas Stuhlmüller. 2013. Knowledge and implicature: Modeling language understanding as social cognition. *Topics in Cognitive Science* 5:173–184.

Grace, Janet, and Paul F. Malloy. 2001. *Frontal systems behavior scale: Professional manual*. Lutz, FL: Psychological Assessment Resources, Inc.

Grace, Janet, Julie C. Stout, and Paul F. Malloy. 1999. Assessing frontal lobe behavioral syndromes with the frontal lobe personality scale. *Assessment* 6:269–284.

Grice, H. P. 1975. Logic and conversation. In *Syntax and semantics*, ed. P. Cole and J. J. Morgan, volume 3, 41–58. New York: Academic Press.

Grodner, D., M. Dalini, S. Pearlstein-Levy, and A. Ward. 2012. Factors that contribute to the use of perspective in referent identification. Talk presented at the 25th annual CUNY Sentence Processing Conference (New York, NY).

Grodner, Daniel J., Natalie M. Klein, Kathleen M. Carbary, and Michael K. Tanenhaus. 2010. Some, and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. *Cognition* 116:42–55.

Happé, Francesca G.E. 1993. Communicative competence and theory of mind in autism: A test of relevance theory. *Cognition* 48:101–119.

Hartshorne, Joshua K., and Jesse Snedeker. submitted. The speed of inference: Evidence against rapid use of context in calculation of scalar implicatures.

Hirschberg, Julia Linn Bell. 1985. A theory of scalar implicature. Doctoral Dissertation, University of Pennsylvania.

Hodges, John R., Naida Graham, and Karalyn Patterson. 1995. Charting the progression in semantic dementia: Implications for the organisation of semantic memory. *Memory* 3:463–495.

Horn, Laurence. 1984. Towards a new taxonomy for pragmatic inference: Q-based and R-based implicatures. In *Meaning, form, and use in context*, ed. D. Schiffrin. Washington: Georgetown University Press.

Howard, D., and K. Patterson. 1992. The pyramid and palm trees test: A test of semantic access from words and pictures. *Bury St Edmunds: Thames Valley Test Company* .

Huang, Yi Ting, and Jesse Snedeker. 2009a. Online interpretation of scalar quantifiers: Insight into the semantics–pragmatics interface. *Cognitive Psychology* 58:376–415.

Huang, Yi Ting, and Jesse Snedeker. 2009b. Semantic meaning and pragmatic interpretation in 5-year-olds: Evidence from real-time spoken language comprehension. *Developmental Psychology* 45:1723–39.

Lafleche, Ginette, and Marilyn S. Albert. 1995. Executive function deficits in mild Alzheimer's disease. *Neuropsychology* 9:313.

Levinson, Stephen. 2000. *Presumptive meanings: The theory of generalized conversational implicature*. Cambridge, Mass: The MIT Press.

Levinson, Stephen C. 1983. *Pragmatics*. Cambridge University Press.

Marty, Paul, Emmanuel Chemla, and Benjamin Spector. 2013. Interpreting numerals and scalar items under memory load. *Lingua* 133:152–163.

Morzycki, Marcin. 2012. Adjectival extremeness: Degree modification and contextually restricted scales. *Natural Language & Linguistic Theory* 30:567–609.

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

Nieuwland, Mante S., Tali Ditman, and Gina R. Kuperberg. 2010. On the incrementality of pragmatic processing: An erp investigation of informativeness and pragmatic abilities. *Journal of Memory and Language* 63:324–346.

Noveck, Ira A., and Andres Posada. 2003. Characterizing the time course of an implicature: An evoked potentials study. *Brain and Language* 85:203–210.

Papafragou, Anna, and Julien Musolino. 2003. Scalar implicatures: Experiments at the semantics–pragmatics interface. *Cognition* 86:253–282.

Paul, Rhea, and Donald J. Cohen. 1985. Comprehension of indirect requests in adults with autistic disorders and mental retardation. *Journal of Speech, Language and Hearing Research* 28:475–9.

Paula McKay, A., Peter J. McKenna, Peter Bentham, Ann M. Mortimer, Alison Holbery, and John R. Hodges. 1996. Semantic memory is impaired in schizophrenia. *Biological Psychiatry* 39:929–937.

Pijnacker, Judith, Peter Hagoort, Jan Buitelaar, Jan-Pieter Teunisse, and Bart Geurts. 2009. Pragmatic inferences in high-functioning adults with autism and asperger syndrome. *Journal of Autism and Developmental Disorders* 39:607–618.

Schulz, Katrin, and Robert Van Rooij. 2006. Pragmatic meaning and non-monotonic reasoning: The case of exhaustive interpretation. *Linguistics and Philosophy* 29:205–250.

Sikos, L., S. Tomlinson, H. Traut, and D. Grodner. 2013. Incremental computation of scalar implicatures: An ERP study. Poster presented at the 26th annual CUNY Sentence Processing Conference (Columbia, SC).

Sperber, Dan, and Deirdre Wilson. 1986. *Relevance: Communication and cognition*. Oxford: Basil Blackwell.

Stout, Julie C, Rebecca E Ready, Janet Grace, Paul F. Malloy, and Jane S Paulsen. 2003. Factor analysis of the frontal systems behavior scale (FrSBe). *Assessment* 10:79–85.

Surian, Luca. 1996. Are children with autism deaf to Gricean maxims? *Cognitive Neuropsychiatry* 1:55–72.

van Tiel, Bob, Emiel van Miltenburg, Natalia Zevakhina, and Bart Geurts. unpublished m.s. Scalar diversity.

Van Rooij, Robert, and Katrin Schulz. 2004. Exhaustive interpretation of complex sentences. *Journal of Logic, Language and Information* 13:491–519.

Wang, A. Ting, Susan S. Lee, Marian Sigman, and Mirella Dapretto. 2006. Neural basis of irony comprehension in children with autism: The role of prosody and context. *Brain* 129:932–943.

Warrington, Elizabeth K., Pat McKenna, and Lisa Orpwood. 1998. Single word comprehension: A concrete and abstract word synonym test. *Neuropsychological Rehabilitation: An International Journal* 8:143–154.

Wilson, Barbara A., Jonathan J. Evans, Hazel Emslie, Nick Alderman, and Paul Burgess. 1998. The development of an ecologically valid test for assessing patients with a dysexecutive syndrome. *Neuropsychological Rehabilitation* 8:213–228.

Xiang, Ming, Julian Grove, and Anastasia Giannakidou. 2013. Dependency-dependent interference: NPI interference, agreement attraction, and global pragmatic inferences. *Frontiers in Psychology* 4:708.

Affiliation

E. Matthew Husband
St. Hugh's College
University of Oxford
matthew.husband@ling-phil.ox.ac.uk