Available online at www.sciencedirect.com

ScienceDirect

Journal of Pragmatics 99 (2016) 78–95

journal of PRAGMATICS

www.elsevier.com/locate/pragma

Why only some adults reject under-informative utterances

Kyriakos Antoniou [a,c,*], Chris Cummins [a,b], Napoleon Katsos [a]

[a] *Department of Theoretical and Applied Linguistics, University of Cambridge, United Kingdom*
[b] *Department of Linguistics and English Language, University of Edinburgh, United Kingdom*
[c] *Département de Langues et Lettres/LaDisco, Université libre de Bruxelles, Belgium*

Received 15 July 2015; received in revised form 29 April 2016; accepted 2 May 2016
Available online 31 May 2016

Abstract

Several studies have investigated how listeners generate scalar implicatures using the under-informative statement paradigm, where participants evaluate statements such as ''Some of the cards have a star'' as descriptions of situations in which all of the cards have a star. Rejection of the under-informative utterances is taken as evidence that participants have interpreted these sentences with a scalar implicature, to the effect that ''Some but not all of the cards have a star''. However, acceptance rates of under-informative utterances exceed 35% in many studies (Bott and Noveck, 2004; Guasti et al., 2005; Pouscoulous et al., 2007; i.a.). The aim of our experimental investigation is to examine the cognitive or personality profile of participants who reject under-informative utterances. We provide empirical evidence that age and working memory capacity significantly predict the rate at which under-informative utterances are rejected, but find little support for influence from a broad range of personality factors.

Keywords: Scalar implicature; Pragmatics; Executive functions; Working memory; Personality factors

1. Introduction

In the Gricean account of implicature (Grice, 1989) a broad category of implied meanings arises from violations of the first maxim of quantity, which enjoins interlocutors to make their contributions ''as informative as is required (for the current purposes of the exchange)'' (Grice, 1989:26). These include the much-discussed case of scalar implicatures (SIs). According to an influential account proposed by Horn (1972), SIs are computed on the basis of pre-existing linguistic scales which order lexical terms (such as ''some'', ''all'') with respect to the strength of the information that they convey. The use of a proposition with a less informative term (e.g. ''some'') implicates that the proposition with the more informative term (e.g. ''all'') does not hold, as in (1)–(2) below.

(1) A: Did all of his students fail the exam?
B: Some of his students failed the exam

(2) Not all of his students failed the exam.

The precise mechanism by which SIs are generated has been the subject of much linguistic debate (Carston, 1998; Chierchia, 2004; Chierchia et al., 2011; Hirschberg, 1991; Geurts, 2010; Levinson, 2000; Sperber and Wilson, 1986/1995;

* Corresponding author at: English Faculty Building, 9 West Road, Cambridge CB3 9DP, United Kingdom. Tel.: +44 7787207681.
E-mail address: ka353@cam.ac.uk (K. Antoniou).

http://dx.doi.org/10.1016/j.pragma.2016.05.001

among others). On the Gricean view, computing a SI is a reasoning process about the speaker's intentions that involves taking into consideration a rich array of linguistic and extra-linguistic information. This includes (a) the literal meaning of the utterance; (b) the purpose of the utterance and assumptions about the speaker's cooperativity; (c) sensitivity to the first maxim of quantity, i.e. awareness that there is a more informative proposition that could have been used but was not; and (d) the interlocutor's epistemic state, i.e. the assumption that the speaker is knowledgeable about the situation. Similar accounts have been proposed by Hirschberg (1991), Geurts (2010), and Relevance theorists (Carston, 1998; Sperber and Wilson, 1986/1995). We call these accounts collectively the ''contextual accounts'' because, even though they may differ in the details of the implicature derivation, they all assume that SIs (a) are always generated through a nonce inferential process of the type proposed by Grice and (b) are context dependent and arise only when certain contextual information is available to the interlocutor (such as the information in b-d in the Gricean derivation above). Horn's (2004, 2005) account of SI might also be classified as a contextual one, at least in terms of the processes that lead to the generation of the SI. In his view, scalar terms normally and most commonly appear in contexts that licence a SI (''default'' contexts), but SI computation is still considered to be a context-driven process.

An alternative proposal is that SIs are generated by default (Chierchia, 2004; Levinson, 2000). According to this view, the scalar inference is always retrieved upon encountering a scalar trigger without reference to the communicative context and without the elaborate reasoning proposed by Grice. Subsequently, the SI may be cancelled if the appropriate contextual assumptions are not met (such as the assumptions in b-d above).

Finally, the grammatical account (e.g. Chierchia et al., 2011; see also Fox, 2007) posits that SI computation is achieved through a covert focus operator **O** which is assigned by the grammar, can take scope over any constituent with a propositional meaning, and has similar properties to the word ''only''. In (1), for instance, this proposal suggests that the silent grammatical operator **O** takes scope over the sentence with the scalar term ''some'' which leads to the negation of the alternative proposition with ''all'' and, hence, to the computation of the SI.

A prolific strand of research has tried to adjudicate between theoretical accounts of SI (particularly, the contextual theory and Levinson's (2000) default theory) using the ''under-informative statement task'' (Bott et al., 2012; Bott and Noveck, 2004; De Neys and Schaeken, 2007; Noveck and Posada, 2003; i.a.). Typically, in this paradigm participants are asked to perform a timed binary truth-value judgement task on sentences such as (3).

(3) Some elephants have trunks.

Rejection of (3) is assumed to indicate the generation of the SI ''some but not all'', whereas acceptance of (3) an interpretation without the SI. Consequently, by comparing rejection and acceptance times, it is possible to compare interpretations with and without the SI, respectively.

Most of these investigations have largely focused on the time-course with which under-informative sentences are rejected as compared to their acceptance. Processing models inspired by the contextual view often assume that SIs incur an additional processing cost compared to semantic meaning (e.g. Bott et al., 2012; Bott and Noveck, 2004; Breheny et al., 2006). This is because computing a SI depends upon contextual information, which is not required for accessing the plain meaning of scalar terms. A processing instantiation of Levinson's (2000) default theory (henceforth, the ''default model''), on the other hand, predicts no processing cost for SIs, since the inference is automatically generated at the lexical level and is relatively context independent.[1] However, the default model further assumes additional costs whenever the SI is not generated. In these cases, the inference, which had been generated by default, is cancelled, leading to processing costs associated with backtracking and re-analysis (e.g. Bezuidenhout and Cutting, 2002). All in all, the majority of studies that employed the under-informative statement paradigm have reported longer response times in the rejection case, which has been taken as evidence in favour of the contextual over the default model, in that SI interpretations appear to be associated with a processing cost (e.g. Bott et al., 2012; Bott and Noveck, 2004; Noveck and Posada, 2003).

However, one finding that has not received much attention within this body of research is that there is always a group of adults who systematically fail to reject under-informative utterances (Bott and Noveck, 2004; Guasti et al., 2005; Pouscoulous et al., 2007; i.a.). This seems to suggest that some (otherwise cognitively and linguistically normal) adults do not derive SIs in response to these stimuli. In this paper we follow up on studies that used the under-informative statement paradigm (Bott et al., 2012; Bott and Noveck, 2004; De Neys and Schaeken, 2007; Noveck and Posada, 2003) in order to examine which personality traits or cognitive factors influence whether adults will reject an under-informative utterance or not.

[1] See Levinson (2000:5 and 104) where he suggests that SIs are derived at the lexical level, without being affected by contextual information or background assumptions; and Levinson (2000:27–29) where he argues that one of the reasons for the existence of default generalised conversational implicatures (like SIs) is that they are ''cheap'' inferences, which maximise the efficiency and speed of communication (see also p. 382, where he explains ''cheap'' in terms of processing time).

1.1. Only some adults reject under-informative sentences

Attempts to explain the observed variation among adults' responses to under-informative statements can be divided into two camps, one placing emphasis on participants' cognitive resources and the other on participants' personality traits.

In the cognitive camp, there are those who suggest that computing a SI is resource demanding (e.g. Bott and Noveck, 2004; De Neys and Schaeken, 2007; Marty and Chemla, 2013). According to contextual accounts, for instance, there are various aspects of SI derivation that are potentially dependent upon cognitive resources: taking the speaker's knowledge state into account, overriding an initial literal interpretation before actually computing the implicature (if one adopts Grice's view) or even coordinating all the linguistic and contextual information that one needs to take into account in order to generate a SI. Similarly, from the perspective of the grammatical account, one might argue that the application of the silent operator **O** is costly and resource consuming (see Marty and Chemla, 2013). There is no direct prediction for why SI computation might consume cognitive resources if one adopts the default model, given that in this account SIs are considered to be lexicalised and to occur automatically and effortlessly.[2] Generally, on accounts that assume SI computation requires processing resources, the failure of some participants to reject under-informative sentences could be attributed to their inability to recruit those processing resources (for whatever reason).

De Neys and Schaeken (2007; see also Dieussaert et al., 2011) were the first to provide experimental evidence that the processing of SIs draws on working memory resources. They reported that burdening adults' working memory with the requirement to remember a complex dot pattern decreased the rate of SI responses to under-informative utterances but did not affect responses to semantically true and pragmatically felicitous sentences. Marty and Chemla (2013) and Marty et al. (2013) further showed that this working memory effect was not due to the process of verifying a SI interpretation. It has been argued, for instance, that in order to verify a SI interpretation of (3) the hearer has to determine two things: whether there are elephants that have trunks and whether there are elephants that do not have trunks (e.g. Grodner et al., 2010). Thus, it is possible that working memory is involved in the verification process of a SI interpretation and not in the inferential process of actually deriving the implicature. To test this, Marty and Chemla (2013) and Marty et al. (2013) used a dual-task methodology (as in De Neys and Schaeken, 2007), but compared the effect of the secondary task on SI responses to under-informative ''some'' sentences and responses to similar ''only some'' sentences. Crucially, the hearer requires access to precisely the same information in order to verify the ''only some'' sentences as they would to verify the ''some'' sentences under a SI interpretation. The results of the Marty et al. (2013) study indicated a working memory effect only on SI responses to under-informative ''some'' sentences.

Nevertheless, it should be noted that none of these studies were specifically designed to test which individual differences contribute to the ability to generate SIs. They merely showed that the rate of implicature generation (specific to a participant) decreases if the participant's working memory (WM) is loaded with some other task. No other cognitive variables were measured, so this research leaves open the possibility that the observed effect of WM might alternatively be explained in terms of ''third factors'', i.e. other cognitive variables that correlate with WM or pragmatic ability.

It is also important to note that the paradigm used in these studies may have actually boosted the importance of WM. In experiments using encyclopaedically under-informative utterances (such as (3)) participants have to search their encyclopaedic memory in order to evaluate the more informative alternative proposition. Additionally, in these studies there might be costs (in memory or processing) from the process of verifying a SI interpretation. Although Marty and Chemla's (2013) and Marty et al.'s (2013) results suggest that WM is not implicated in the verification process, this evidence can be challenged. Specifically, Tomlinson et al. (2013) have recently argued that while the bare quantifier ''some'' focuses attention on the referent set (e.g. elephants that have trunks), ''only some'' focuses attention on the complement set (e.g. elephants that do not have trunks). Thus, in these studies it is possible that cognitive load did not interact with participants' responses to ''only some'' sentences because the focusing properties of ''only'' made it easier for participants to verify and reject these statements. In short, two important questions remain open: is the effect of WM reported in previous studies specifically associated with the process of computing a SI? Moreover, if so, does WM still have an integral role in this process when the memory load of evaluating the more informative alternative is minimal?

Taking a categorically different approach, personality-based accounts propose that personality traits might be the source of variation in how adults interpret under-informative statements. According to Katsos and Bishop (2011), a response to an under-informative statement (either acceptance or rejection) is based on a metalinguistic decision, which is taken independently of the participants' pragmatic knowledge. Specifically, they hypothesise that this decision may be affected by personality factors such as pedantry.

Similarly, Feeney and Bonnefon (2013) provided experimental evidence that personality characteristics affect the interpretation of scalar terms. They asked participants to interpret sentences using the scalar term ''or'' and found that

[2] Note, though, that other instantiations of the default view, besides Levinson's (2000), might also predict a cognitive cost for SI interpretation (see Section 4).

adults were more likely to interpret ''or'' logically (i.e. compatible with ''and'') when one character announced to another that two undesirable events would happen. This is because, in this context, it is possible to think that the speaker knew that in fact both events would happen, but for reasons of politeness used the disjunctive ''or''. They also showed that participants' self-rated honesty was associated with the propensity to interpret ''or'' with a SI (i.e. ''not both''), a finding consistent with the claims of personality-based accounts. They argued that this is because participants who see themselves as honest are more committed (and expect others to be more committed) to honesty and maximal informativeness at the expense of politeness considerations.

In a similar vein, Nieuwland et al. (2010) reported an effect of the prevalence of autistic traits on the processing of under-informative sentences. They presented participants with under-informative utterances such as ''Some people have lungs'' while recording their N400 effect modulation (an electrophysiological index of how well the semantic meaning of a word fits the linguistic context) at the critical word in the end of these utterances. They found that normal participants with a stronger tendency towards autistic traits were less sensitive to under-informativeness (i.e. they exhibited a smaller N400 effect).

To summarise, currently there are two broad accounts that try to explain why adults vary in their responses to under-informative utterances. On the one hand, some researchers assume that this variation is caused by differences in cognitive resources. On the other hand, it has been argued that the split between responders can largely be explained by differences in personality traits, whether in terms of pedantry, expectations of honesty, or prevalence of autistic traits. Our experiment is the first direct test of the relative contributions of the two types of influences.

2. Experiment

The principal aim of our experiment was to investigate which cognitive or personality factors predict whether or not an adult will reject an under-informative utterance. All participants were given an under-informative statement task and an extensive battery of cognitive and personality tests relevant to SI generation. Participants' age, gender, non-verbal fluid intelligence, and verbal intelligence were also measured as potential confounding variables.

In the under-informative statement task, participants performed truth-value judgments on sentences using the quantifiers ''some'', ''all'', and ''none'' as descriptions of visual displays. The displays showed five cards depicting various objects. There were three types of sentences: true and informative, semantically false, and under-informative. Critical sentences were of the under-informative type using the quantifier ''some'' (see Appendices A and B).

Unlike most prior studies, we tested the (in)felicity of under-informativeness using visual displays rather than matters of encyclopaedic knowledge. This has several advantages. First, as pointed out by Guasti et al. (2005), when presented with de-contextualised statements, participants are free to conjure up their own contexts. Thus, participants might accept putatively under-informative utterances because they include atypical tokens in their context (e.g. participants who accept (3) may think of maimed elephants). This problem is avoided when the context of evaluation is delineated by the experimenters. Secondly, as argued in the previous section, verifying a SI interpretation for an encyclopaedically under-informative utterance may involve extra cost (in (3), for instance, the hearer needs to establish whether elephants without trunks might exist). The use of visual displays ensures that all the information required for verifying an interpretation is readily available in the visual context. This circumvents the possibility of memory costs associated, not with the process of generating a SI, but with the process of searching memory in order to verify a SI interpretation.

2.1. Method

2.1.1. Participants

63 adult participants (ages 18–46, mean age 24, SD 5.9 years, 46 female) were recruited by advertising within the University of Cambridge.

2.1.2. Materials and procedure

All tests were taken in the laboratory with the exception of the Autism-Spectrum Quotient (AQ), the Big Five Inventory (B5) and the Honesty/Integrity/Authenticity scale (HIA). These were completed before the participant came to the laboratory. Thirty-nine participants were tested in two sessions of approximately 30 min each, four participants showed up only for the first session, and 21 participants took all tests in a single session. The fixed order of tasks in the first session was: Under-informative statement task, Sentence recall task, Backward Digit Span task (BDST), and WASI matrix reasoning test (WASI). The fixed order of tasks in the second session was: Simon task, Reading Span Task, Stroop task, and Number-letter task. Participants tested in a single session took the tests in the following order: Under-informative statement task, Simon task, Reading Span Task, Sentence recall task, BDST, Stroop task, Number-letter task, and WASI.

2.1.2.1. Under-informative statement task. In each trial of this test, the participant saw a depiction of five cards face down. A fixation cross was shown (500 ms) and then an auditory stimulus was played, ''There are <X> on <Q> of the cards'', where X was the item type (rings, hearts, squares, stars, suns, or moons) and Q the quantifier (''all'', ''some'', or ''none''). When the auditory stimulus ended, the cards were immediately ''turned over'' to reveal the items. Participants were instructed to press a green-labelled key if the utterance were true and a red-labelled key if it were false, responding as quickly and accurately as possible.

The experiment was implemented using E-Prime. Forty-two trials were administered to each participant, comprising six blocks of seven trials (see Appendices A and B). There were six under-informative items using the quantifier ''some'', each presented in one of the six blocks. The auditory stimuli were recorded by a native English speaker. They were pronounced without any contrastive focus accent on the quantifier, matched for length, and constructed from the same recordings. The items were introduced in a set of two practice trials, neither of which were repeated in the experiment.

2.1.2.2. Working memory tests

2.1.2.2.1. The Backward Digit Span Task (Wechsler, 1944). In each trial, participants heard an auditory stimulus presenting a list of numbers (e.g. ''5'', ''6'', ''3''), and were asked to repeat the numbers in reverse order. The auditory stimuli were recorded by a native English speaker. Each stimulus was embedded in a Microsoft PowerPoint slide. The task began with three digits and the string of digits increased after two trials on the same number of digits. It was discontinued when participants erred on two trials on the same number of digits (the highest level was eight digits). There were two practice trials with three digits each. One point was awarded for each trial where all numbers were recalled in the correct order. An experimenter scored participants' responses on a scoring sheet with the correct answers for each trial. Split-half reliabilities for this test for the age range tested are from .66 to .71 (Wechsler, 1944).

2.1.2.2.2. The Reading Span Task (Kane et al., 2004). This task was a version of the Reading Span Task developed by Kane et al. (2004) on E-Prime software. In each trial participants were presented with a series of sentences. They were instructed to judge whether the sentences made sense or not and to remember a capitalised letter at the end of each sentence. At the end of each trial, participants were presented with a recall cue (three question marks) and had to write down each letter from the preceding set of sentences in the correct order. There were 12 trials and in each trial the set of sentences ranged from two to five. There were three practice trials with three sentences each. Participants were given a scoring sheet on which to write down their responses. Participants were given one point for each correct letter recalled in the correct order (see Unsworth et al., 2009). For this task Conway et al. (2005) reported internal consistencies that range from .70 to .79 depending on the scoring method used. Unsworth et al. (2009) used an automated, slightly different, version of this task that employed the exact same scoring procedure. They reported internal consistency of .89 (Cronbach's α) and test-retest reliability of .82.

2.1.2.3. Inhibitory control tests

2.1.2.3.1. The Stroop Task (Stroop, 1935). This version of the Stroop task was implemented on E-Prime software. It consisted of three blocks, each including 48 trials (eight practice trials) and presented in a fixed order. In the first block, participants saw strings of Xs (e.g. ''XXX'') in different font colours. In the second block colour words (e.g. ''red'') were presented in a congruent colour (e.g. ''red'' in red font) and in the third block they were displayed in incongruent colours (e.g. ''red'' in blue font). The colours red, green, blue, purple, and yellow were used. Five keys on the keyboard were colour-labelled and participants were instructed to indicate the font colour of each word by pressing the corresponding key on the keyboard as quickly as possible. Successful performance in incongruent trials required inhibition of the tendency to actually read the colour name. Accuracy and reaction times (RTs) were recorded. The dependent measure was the difference in RTs between the congruent and incongruent conditions.

2.1.2.3.2. The Simon Task (Simon, 1969). In this task, implemented using E-Prime, participants were instructed to press the right arrow key on the keyboard if a red square appeared on the screen and the left arrow key if a green square appeared. In congruent trials, each square appeared on the same side as the correct button (e.g. a red square on the right of the screen). In incongruent trials, the square appeared on the side opposite to the appropriate arrow key (e.g. a red square on the left of the screen). This created a conflict between an irrelevant feature of the stimulus that had to be inhibited (the square's position on the screen) and a relevant feature that should have determined the correct response (the square's colour). In neutral trials, the square appeared on the centre of the screen. The task included two blocks of trials presented in a fixed order. In the first block, 48 congruent and 48 incongruent trials were randomly intermixed. The second block included only 48 neutral trials. In both blocks, the test trials were preceded by eight practice trials. Accuracy and RTs were recorded. The dependent measure was the difference in RTs between congruent and incongruent trials in the first block.

2.1.2.4. Cognitive flexibility test

2.1.2.4.1. The Number-letter task (Miyake et al., 2000). This was a version of the task used by Miyake et al. (2000). In this task a number-letter pair (e.g. ''7G'') appeared in one of two squares at the top or bottom part of a grid presented on the centre of the computer screen. There were three blocks of trials presented in a fixed order. The first two blocks required no switching and consisted of 40 trials each (eight practice trials). The third block included 136 trials (eight practice trials), 64 of which required switching. In the first block, a number-letter pair appeared in the two top squares of the grid and participants were asked to indicate whether the number was odd or even by pressing a green- or a red-labelled key on the keyboard, respectively. In the second block, the Number-letter pair appeared in the bottom squares and participants were required to indicate whether the letter was a vowel or a consonant, again by pressing the green- or the red-labelled key, respectively. Finally, in the third block the Number-letter pair appeared in one of the four squares in a clockwise fashion: participants were instructed to respond as in the first block when the characters were in the top squares and to respond as in the second block when the characters were in the bottom squares. Thus, in half of these trials, participants had to switch between categorising the number and categorising the letter. Participants were instructed to respond as fast and as accurately as possible. Accuracy and RT were recorded. The dependent variable was the RT difference between switch trials in the third block and repeat trials in the first two blocks. For this task, Miyake et al. (2000) report a split-half reliability of .91.

2.1.2.5. Test of autistic traits

2.1.2.5.1. The Autism-Spectrum Quotient (Baron-Cohen et al., 2001). This questionnaire measures the degree to which adults show autistic-like traits in their everyday behaviour. It comprises fifty questions, covering five areas known to be associated with the autism spectrum: social skill, attention switching, attention to detail, communication, and imagination. Autistic behaviour is associated with poor social, communication, and imagination skills, excessive attention to detail, and strong focus of attention. Higher scores indicate stronger levels of autistic traits. In our study, the AQ was completed online using the free internet software Qualtrics (http://qtrial.qualtrics.com). Test–retest reliability for the whole questionnaire is .70 (Baron-Cohen et al., 2001). Measures of internal consistency (Cronbach's α) are as follows: communication = .65, social = .77, imagination = .65, attention to detail = .63, and attention switching = .67 (Baron-Cohen et al., 2001).

2.1.2.6. Tests of personality traits

2.1.2.6.1. The Big Five Inventory (John et al., 2008). The B5 is an instrument measuring the prototypical components of five dimensions of personality: extraversion, agreeableness, conscientiousness, neuroticism, and openness. It contains 44 items and the items in each scale form the core definition of each of the five factors. Extraversion includes traits such as being energetic, sociable, adventurous, and enthusiastic. Agreeableness covers facets such as being sympathetic, helpful, forgiving, and cooperative. Conscientiousness captures traits such as being organised, thorough, cautious, and responsible. Neuroticism contains characteristics such as getting easily anxious, tense, being unstable, and being emotional. Openness measures facets such as being imaginative, intelligent, artistic, and curious. The B5 was also taken online using Qualtrics (http://qtrial.qualtrics.com). Alpha reliabilities for the five scales range from .75 to .90. Test-retest reliabilities range from .80 to .90 (John et al., 2008).

2.1.2.6.2. Honesty/Integrity/Authenticity Scale (Goldberg et al., 2006; http://ipip.ori.org). This is a brief questionnaire consisting of nine items (two of which are fillers). It was administered as a measure of participants' self-rated honesty. The scale included items such as ''I believe that honesty is the basis for trust'' or ''I lie to get myself out of trouble''. It was administered online using Qualtrics (http://qtrial.qualtrics.com). It has an internal reliability coefficient of .72 (http://ipip.ori.org).

2.1.2.7. Test of non-verbal IQ

2.1.2.7.1. The WASI matrix reasoning test (Wechsler, 1999). In each item of this test, participants had to complete a matrix from which a section was missing by selecting one of five possible choices. Split-half reliabilities range from .88 to .96. Test-retest reliability is .72 for the age range in this sample (Wechsler, 1999).

2.1.2.8. Test of verbal IQ

2.1.2.8.1. The sentence repetition task. This task comprised 17 sentences (one for practice) of increasing length and syntactic complexity. The items were based on the NEPSY sentence repetition battery (Korkman et al., 1999), but were modified by the authors in order to remove the easiest items (most suitable for children) and replace them with more complex sentences (see Appendix C). In each trial, participants heard a pre-recorded sentence spoken by a native

English speaker. Each stimulus was presented orally in a Microsoft PowerPoint slide. Participants were asked to listen to each sentence carefully and repeat the sentence exactly as heard. An experimenter scored participants' responses on a scoring sheet with the full sentences for each trial. Two points were awarded for each perfectly recalled sentence, one point for one or two errors, and no points for more than two errors.

3. Results

3.1. Rate of SI responses to the critical items of the under-informative statement task

For the critical under-informative items, 43 participants responded true more than three times out of six, 18 responded false more than three times, and two accepted them exactly half of the time. Fig. 1 displays the distribution of participants as a function of the number of times they rejected the critical items (0 up to 6).

Significance tests of skewness and kurtosis showed that our dependent measure, the rate of rejection of critical items (Uinftotal), violated the assumption of normality (*z*-score skewness: 0.895/0.30 = 2.98, $p < 0.01$). Thus, we performed all correlational analyses using the bootstrap method, which is accurate even when assumptions are violated. All bootstrap analyses were conducted in SPSS requesting 1000 bootstrap samples and bias-corrected and accelerated bootstrap confidence intervals instead of the default percentile intervals.

3.2. Personality and cognitive measures

Reliability statistics for all tests, for the current sample, are presented in Table 1.[3] For each task, we estimated the split-half reliability using the Spearman–Brown correction.

We also created composite scores for all variables that were conceptually related. Thus, the two working memory (BDST and Reading Span Task) and the two inhibition measures (Simon and Stroop tasks) were collapsed into a single WM and a single inhibition score, respectively. Missing data was substituted using the mean for each variable. This procedure affected 3% of the data in the WASI, 6% of the data in the HIA, 6% of the data for the Inhibition composite score, and 11% of the data in the Number-letter task.

3.3. The relation between the personality and cognitive measures and the rate of rejection of under-informative items (Uinftotal)

In order to explore the relations between the factors tested and the dependent measure of interest (Uinftotal), we first looked at the bivariate and partial correlations. These analyses were exploratory in nature, on the grounds that we had too many variables (19) for the participant sample ($n = 63$) to directly conduct a regression analysis. When looking at the bivariate correlations, only age significantly and negatively correlated with our dependent measure (*r*(two-tailed) = −.354, bootstrap 95% CI [−.502, −.185]). The partial correlations, on the other hand, showed that, when the effect of age, gender, verbal abilities, and IQ was controlled for, only WM significantly and positively correlated with the dependent measure (*r* (two-tailed) = .243, bootstrap 95% CI [.026, .443]).

In order to further explore these relations, a regression analysis was conducted. The analysis included only the two variables that significantly correlated with the Uinftotal measure (age and WM). We also included the variables that significantly correlated with these two variables in order to partial out the shared variance between these variables. These were the switch cost from the Number-letter task for age (*r*(bivariate, two-tailed) = .25, bootstrap 95% CI [.011, .454]) and the WASI score (*r*(bivariate, two-tailed) = .43, bootstrap 95% CI [.226, .602]) and the Openness score from the B5 (*r*(partial, two-tailed) = −.311, bootstrap 95% CI [−.505, −.109]) for WM. We used the forced entry method (with all variables included in a single block) and the bootstrap robust procedure in SPSS.

[3] Reliabilities for the individual tests were estimated as follows: for the Simon, Stroop, Number-letter, BDST, and WASI matrix reasoning tasks, we computed two sub-scores for each participant. For the Simon task sub-score A was the difference in mean RTs between the first 50% of the incongruent trials and the first 50% of the congruent trials in block 1. Sub-score B was the difference in mean RTs between the second 50% of the incongruent trials and the second 50% of the congruent trials in block 1. For the Stroop task sub-score A was the difference in mean RTs between the first 50% of trials in block 3 (incongruent condition) and the first 50% of trials in block 2 (congruent condition). The second sub-score was the difference in mean RTs between the second 50% of the trials in block 3 (incongruent condition) and the second 50% of the trials in block 2 (congruent condition). For the Number-letter task, sub-score A was the difference in mean RTs between the first half of the switch trials in block 3 (switch condition) and the trials in block 1 (repeat condition). Sub-score B was the difference in mean RTs between the second half of the switch trials in block 3 (switch condition) and the trials in block 2 (repeat condition). For the BDST and WASI matrix reasoning test, sub-score A was the sum of correct odd trials and sub-score B was the sum of correct even trials. For the rest of the tasks, the split-half reliability was calculated based on all items of each task.

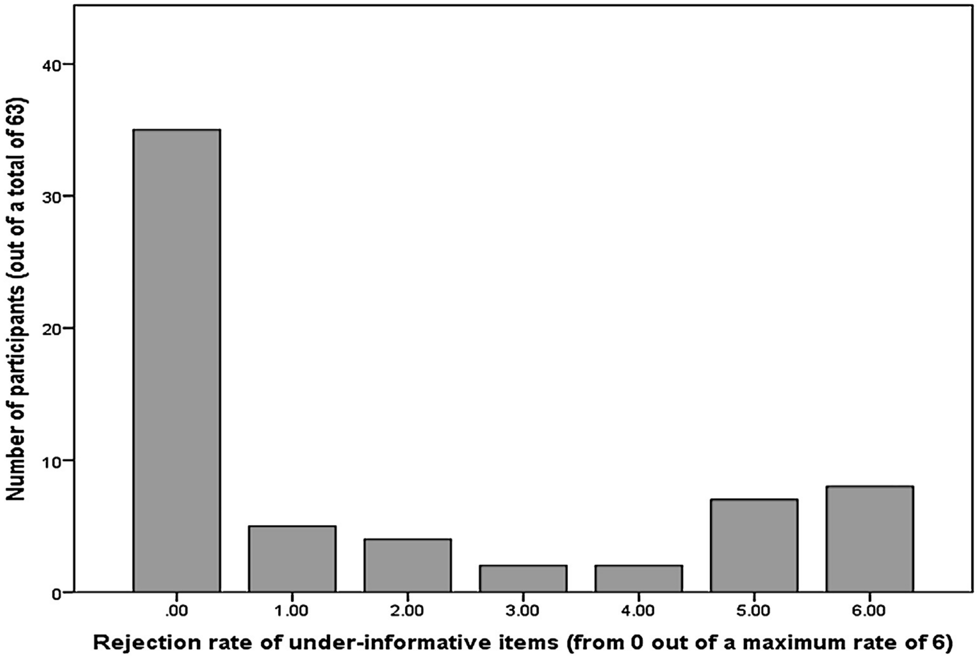


Fig. 1. Distribution of participants depending on the rate of rejection of under-informative items.

Table 1
Split-half reliabilities (using the Spearman–Brown correction) for all tests for the sample tested.

WASI matrix reasoning test	0.503
Sentence repetition task	0.616
Honesty Integrity Authenticity scale	0.340
Autism Quotient Social skills scale	0.619
Autism Quotient Attention switching scale	0.470
Autism Quotient Attention to detail scale	0.668
Autism Quotient Communication scale	0.612
Autism Quotient Imagination scale	0.346
Big Five Inventory Extroversion scale	0.908
Big Five Inventory Agreeableness scale	0.794
Big Five Inventory Conscientiousness scale	0.702
Big Five Inventory Neuroticism scale	0.863
Big Five Inventory Openness scale	0.811
Simon task	0.452
Stroop task	0.255
Number-letter task	0.890
Reading span task	0.728
Backward digit span task	0.770

Table 2 summarises the results of the multiple regression analysis. Overall, this regression model was marginally significant ($F(5, 57) = 2.335$, $p = .054$) accounting for 17% of the variance in the dependent measure. When looking at the coefficients, age and WM significantly predicted performance in the dependent variable, the correlation with age being negative ($t(57) = -2.24$, $p < .05$, bootstrap estimates) and the correlation with WM being positive ($t(57) = 2.04$, $p < .05$, bootstrap estimates).

Given that the overall model was only marginally significant and in order to further validate the results of the bootstrap multiple regression analysis, we conducted a similar logistic regression analysis. In order to perform this analysis, we divided participants into two groups. We use the term ''Pragmatic responders'' for the group of participants who rejected four or more under-informative utterances (17 participants), and ''Logical responders'' for the rest (46 participants). This dependent measure (Uinfsplit) was used instead of the Uinftotal measure. The results of the logistic regression analysis indicated a marginally significant effect of age and a significant effect of WM (Wald = 3.441, $p = .064$ and Wald = 4.269,

Table 2
Results of bootstrap multiple regression analysis.

	B (SE)[a]	*β*
Constant	8.136 (3.460)	
Age	−.123[*,b] (.055)	−.302
WM	.723[*] (.355)	.247
WASI	−.142 (.117)	−.154
Switch cost	−.001 (.001)	−.079
B5Openness	.027 (.055)	.070

Note 1: R^2 = .17. *F*-test = 2.235 (p = .054[c]). *Note 2*: Age = age in years; WM = working memory composite score; WASI = non-verbal fluid intelligence score; Switch = switch cost score from the Number-letter task; B5Openness = score in the openness sub-scale of the big five inventory.

[*] $p < .05$.

[a] Standard errors are bootstrapped.

[b] Significance tests are based on bootstrap estimates.

[c] The *F*-ratio and corresponding *p* value are for the parametric, non-bootstrap multiple regression analysis. SPSS does not provide bootstrap results for the ANOVA analysis that tests the overall significance of the model.

Table 3
Results of final logistic regression analysis.

		95% CI for odds ratio		
	B (SE)	Lower	Odds ratio	Upper
Constant	7.519 (4.290)			
Age	−.153* (.082)	.731	.858	1.009
WM	1.206* (.584)	1.064	3.342	10.496
WASI	−.165 (.144)	.640	.848	1.123
Switch cost	−.001 (.001)	.997	.999	1.002
B5Openness	.000 (.051)	.905	1.000	1.106

Note 1: R^2 = .166 (Cox & Snell), .241 (Nagelkerke). Model χ^2 = 11.420, $p < .05$. *$p < .05$.
Note 2: Age = age in years, WM = working memory composite score, WASI = non-verbal fluid intelligence score, Switch = switch cost score from the Number-letter task, B5Openness = score in the Openness sub-scale of the big five inventory.

$p < .05$, respectively). In addition, the overall model was significant ($\chi^2(5) = 11.420$, $p < .05$) accounting for 16.6–24.1% of the variance in the dependent measure. Results of the logistic regression analysis are presented in Table 3.

4. Discussion

The main empirical finding of this experiment is that working memory and age are significant predictors of whether a participant will reject an under-informative statement. As regards WM, the finding confirms the predictions of cognitive resources accounts (Bott and Noveck, 2004; De Neys and Schaeken, 2007; Marty and Chemla, 2013; i.a.), while personality-inspired accounts (Feeney and Bonnefon, 2013; Katsos and Bishop, 2011; Nieuwland et al., 2010) receive no support. Turning to the effect of age, whereby older adult participants generate fewer SIs, there is no expectation in the literature for such a finding. Nevertheless, it is possible that the absence of any reported effects of age in the experimental literature is largely due to a lack of focus on this issue. Some studies recruited from a narrow age range (e.g. participants in Dieussaert et al., 2011 were aged 17–19 years) while in many studies age was not factored in as a potential predictor in the statistical analyses (e.g. Bott et al., 2012; Bott and Noveck, 2004). Therefore, it is not surprising that if an effect of age does exist, it would only be captured by an experiment like this one, where the age range was relatively wide (18–46 years), and age was included in the analyses.

4.1. The relation between SI generation and working memory

The finding of a positive relation between working memory capacity and rate of SI generation in our study is directly in line with previous findings by De Neys and Schaeken (2007), Dieussaert et al. (2011), Marty and Chemla (2013), and Marty et al. (2013). Focusing on group performance, the studies by De Neys and Schaeken (2007), Marty and Chemla (2013), and Marty et al. (2013) showed that experimentally limiting the WM capacity of adult participants leads to lower rates of SI responses. Similarly, but taking an individual differences approach, our experiment showed that participants with low WM capacity are less likely to interpret under-informative stimuli with a SI. Thus, the two lines of research seem to converge on the same conclusion: limited WM capacity reduces the likelihood of deriving SIs.

This is not to suggest of course that the cost of deriving a SI is so large that some neuro-typical adults in our study did not have the necessary WM resources to implement the SI interpretation process at all. We believe, instead, that aspects of our experimental design placed additional cognitive demands on our participants, besides those normally required for the calculation of the implicature. This led to additional pressure on participants' WM resources, perhaps analogously to the cognitive load manipulation in previous research, leaving little space in WM for implicature computation and, thus, resulting in fewer pragmatic responses to under-informative sentences. We propose here that this additional, non-pragmatic burden on WM was imposed by the high number of filler items in the under-informative statements task in our study (with a proportion of .86 of filler items) (see also Pouscoulous et al., 2007; De Neys and Schaeken, 2007, for similar arguments).

Some direct support for this view comes from the two experiments conducted by Dieussaert et al. (2011). As already noted earlier, in their first experiment, Dieussaert et al. (2011) employed a dual-task methodology and reported a lower percentage of pragmatic responses under cognitive load. This effect was mostly evident in participants with low WM capacity. In their second experiment, Dieussaert et al. (2011) manipulated the number of filler items in the under-informativeness task in order to examine additional factors that may affect pragmatic responding. One group of participants was presented with a low proportion of filler statements (.67) and a different group with a higher proportion (.83). Their results indicated an overall decrease in consistency of participant responses but also a lower percentage of pragmatic answers to under-informative sentences in the latter group of participants.

In line with the proposal put forth here, Dieussaert et al. (2011:2365) suggest that increasing the number of filler statements had a similar negative influence on SI responses to the cognitive manipulation in their first experiment. Thus, the results of the two experiments by Dieussaert et al. (2011) suggest that just as a secondary memory task places a burden on participants' WM resources and results in lower percentages of pragmatic responses among low-working memory individuals, so does the high number of filler items. It is thus possible that the task demands on participants with low WM in our study were sufficient to obstruct pragmatic processing and the generation of SI responses.

That said, we also suggested that the inferential process of computing a SI per se, also places demands on WM. Our experiment was not specifically designed to test which aspects of SI processing require the involvement of WM resources (or the negative contribution of age, see next section). In this respect, we can only offer speculations about the exact nature of this relation.

First, we should mention that the effect of WM on the rate of rejection of under-informative utterances is incompatible with the default model of SIs as expressed by Levinson (2000). This is because a direct prediction from this account is that SIs are triggered at the lexical level and that they are cost free. Levinson's (2000) account, however, is just one of the possible characterisations of the default view. Another possible instantiation is that SIs are always triggered, irrespective of contextual support, but because some of the pragmatic machinery described by Grice (1989) still takes place (e.g. the activation and comparison of alternative representations, the assumption that the speaker knows what he is talking about), this renders the default implicature costly. Nevertheless, for empirical and conceptual reasons, we do not believe that this instantiation of the default view is on the right track. First, even though this account correctly predicts that SI generation requires WM resources (due to the application of some Gricean-like pragmatic stages), it also predicts (like any other implementation of the default view) that non-implicature interpretations of scalar terms should be even more cognitively demanding (and, hence, have an even stronger relation with WM). This is because of the need to implement an additional step of withdrawing a default implicature. Clearly, the second prediction does not receive any support from our results. Second, the idea of costly defaults is less intuitively and conceptually appealing given that one of the main motivations for positing defaults is that they maximise the efficiency and speed of communication (Levinson, 2000).

Having said that, there are several potential explanations of the relation between WM and SI generation depending on the theoretical account to which one is committed. From the perspective of the grammatical account there have been suggestions in the literature that the application of **O** by the grammar might be costly and potentially dependent upon cognitive resources, such as WM (see Marty and Chemla, 2013).

From the contextual accounts' view several explanations seem plausible. As Relevance Theory (Sperber and Wilson, 1986/1995) suggests, WM resources may be part of the cognitive effort that is inherently required when making pragmatic inferences in general. However, it is also possible that the requirement for these resources is not integral to pragmatic processing in general, but instead is related to specific aspects of SIs. We consider three possible instantiations of this idea.

First, WM resources might be required so that alternative representations of the scalar statement can be compared, as has been proposed in the developmental literature (e.g. Chierchia et al., 2001).[4] There is evidence that children's performance with SIs significantly improves when the alternative representations are made salient by the task (Guasti

[4] Note that this prediction arises not only from the contextual accounts of SI, but from the grammatical account as well.

et al., 2005; Papafragou and Musolino, 2003) or when alternative representations are explicitly given to the children (Chierchia et al., 2001). However, experimental evidence with children and adults suggests that participants are invariably sensitive to the first maxim of quantity when judging under-informative stimuli (Katsos and Bishop, 2011; Katsos and Smith, 2010), which seems to speak against this explanation. Katsos and Bishop (2011) and Katsos and Smith (2010) presented 5–7-year-old children and adult participants with descriptions of situations that were either true and informative, or true but under-informative, or false. One group of participants was asked to judge the utterances as right or wrong, and a different group to rate on a (three- or five-point) Likert scale how well the utterances described the situations. Results indicated that, in the former condition, participants were split into two groups, those who habitually accepted and those who habitually rejected the utterance. In the latter condition, however, all participants penalised the under-informative utterances, rating them lower than the informative but higher than the false utterances. One way to interpret these findings is by making a Gricean distinction (see Section 1) between sensitivity to under-informativeness and actual derivation of a SI (the latter step involving the negation of the more informative statement): even though only some participants generated SIs in the binary-judgement task (only some of them rejected the under-informative statements), all participants were sensitive to under-informativeness in the Likert-scale task (as evidenced by their ratings). This interpretation suggests that it is not the activation and comparison of alternatives that is problematic or taxing for participants when deriving SIs. Detecting that an utterance is under-informative still requires the participant to access an alternative relevant representation and to compare it with the scalar statement used, but all children and adults seem to go through this step without difficulty (as shown by the unanimous penalisation of under-informative stimuli in the Likert-scale experiment).

A second possibility is that WM is required so that an initial semantic interpretation, accessed before computing a SI, can be overridden in the face of incoming contextual information (cf. Tomlinson et al., 2013). This would be analogous to the case of syntactic garden-paths where the hearer is temporarily committed to an initial interpretation that, based on late-arriving syntactic cues, proves to be wrong and has to be revised. Recently, experimental research has provided some evidence that WM is implicated in adults' ability to recover from such syntactic garden-paths (Novick et al., 2013). However, this account would need to be reconciled with most of the experimental literature which has provided abundant evidence against a literal-first derivation of implicatures (e.g. Cacciari and Tabossi, 1988; Giora, 2003; McElree and Nordlie, 1999; Shapiro and Murphy, 1993; but see Huang and Snedeker, 2009; Tomlinson et al., 2013).

A final explanation could be that the contribution of WM is linked to another Gricean component of SI generation. The Gricean account posits that SIs are derived only if the listener can assume that the speaker is knowledgeable of the situation and that s/he is in a position to know that the stronger, more informative term does not hold. Thus, calculating a SI involves a Theory of Mind (ToM) component in terms of keeping track of the interlocutor's epistemic state. Recent experimental evidence shows that this inferential step is psychologically real (Bergen and Grodner, 2012; Breheny et al., 2013) and that implicatures do not go through when the speaker is unreliable (Grodner and Sedivy, 2005). This research is suggestive that the role of WM in SI generation might be related to the requirement to take into account the interlocutor's knowledge state that, according to contextual accounts, is inherent in this process.

If this explanation is on the right track, how, then, does WM relate to SI interpretation vis-à-vis ToM? Experimental research suggests that applying ToM in order to understand other people's linguistic behaviour can be slow and costly, it depends on executive resources (including working memory) and is disrupted under cognitive load (e.g., Apperly et al., 2006; Keysar et al., 2003; Lin et al., 2010; Schneider et al., 2012). This evidence mostly comes from paradigms where there is a clear discrepancy between the participant's and the interlocutor's perspectives and participants possess privileged knowledge. Effortful ToM processing in these experiments possibly stems from the need to hold in memory two contrasting perspectives, to keep their representation separate, to inhibit one's own privileged knowledge, and use this information in the process of understanding language. In a similar vein, when computing SIs, participants need to employ their ToM skills in order to infer the speaker's knowledge state, keep that information in mind and (possibly) combine it with other linguistic and non-linguistic cues in order to reach a pragmatic interpretation. This suggests that the WM involvement in SI processing consists not only in the application of ToM, but also in using the output of the ToM-based reasoning process: holding it in memory and coordinating it with other sources of information.

This proposal further suggests that the relation between WM and SI interpretation primarily involves the executive component of WM, the ''central executive'' in Baddeley and Hitch's (1974) terms. According to the influential model proposed by Baddeley and Hitch (1974; see Baddeley (2000) for an updated version), WM is a multicomponent system that consists of short-term memory systems responsible for the temporary storage of verbal and visuospatial material (the phonological loop and the visuospatial sketchpad, respectively) and a central executive, a complex system that acts upon, manipulates, and controls the information in the other two systems. The suggestion that SI interpretation mainly involves the central executive processes of WM receives support from two further considerations. First, the Backward Digit Span Task and the Reading Span task used in our study are tests that are known to primarily load on the executive part of WM (even though they also tap on the phonological loop). Second, there is relevant research on other types of pragmatic processing. For instance, the processing of metaphors has also been shown to involve WM resources, but the contribution of WM is found only when it is measured with tasks that primarily tap on the executive aspect (such as the Backward Digit

Span Task) and not when tested with tasks that primarily tap on short-term memory (such as recalling digits in forward order) (Chiappe and Chiappe, 2007; Mashal, 2013).

4.2. The relation between SI generation and age

Turning to the negative effect of age, we should first clarify that age was not a variable we were originally interested in testing, given that no linguistic or psycholinguistic theory makes any explicit prediction as to whether and how age might affect the derivation of SIs (at least in adults). Rather, age, along with IQ, verbal IQ, and gender, was included in the analyses as a potentially confounding variable.

How, then, can we account for this unexpected finding? A possibility we cannot rule out is that this effect is related to other factors not tested in this study. From the perspective of personality-based accounts, possibilities include the participants' socioeconomic status, number of years at university, or subject of study at university. From a cognitive perspective, a more interesting possibility arises, specifically that the effect of age is also driven by the participants' ability to use their ToM skills in order to derive SIs. Research with adults suggests that ToM abilities decline with ageing (although this has been shown for adults above 50 years old; see Bernstein et al., 2011; Henry et al., 2012) and that adults can show limitations in their use of this ability (e.g. Apperly et al., 2010; Keysar et al., 2000, 2003). This literature suggests that the effect of age in our study might have been due to older adults being less likely to recruit their ToM skills and generate a SI in response to the critical items.[5]

4.3. The relation between SI generation and personality factors

Our results provided no support for the view that personality traits play a role in how people interpret scalar expressions. First, contrary to what one might have expected based on Nieuwland et al.'s (2010) results, we did not find an effect of autistic-like communication difficulties on adults' interpretation of under-informative sentences. We have two points to make about this issue, however. First, as argued in Section 1, Nieuwland et al.'s (2010) study was not a proper individual differences study in that they did not measure any control variables or any other variables that are potentially relevant to the interpretation of under-informative stimuli. Thus, it is possible that (part of) the effect observed in their study might have been due to other factors that correlate with pragmatic ability and/or autistic behaviour, but were not tested in their experiment. For instance, experimental evidence suggests that autism is associated with a dysfunction of the executive system, which subsumes working memory (although the evidence is controversial for WM and is mostly reported for visual WM; see Hill, 2004; Poirier et al., 2011; Richmond et al., 2013; Schuh and Eigsti, 2012). This leaves open the possibility that (part of) the contribution of autistic traits in Nieuwland et al. (2010) might have been due to WM and that this was confounded with the effect of autistic traits, which would enable our findings to be reconciled with theirs (but see footnote 5).

Another possibility is that the two studies obtained different results because they used different methodologies. The measure in our task was a behavioural measure (rate of rejection of critical stimuli) while Nieuwland et al.'s (2010) was an

[5] An anonymous reviewer raised the possibility that if this explanation is on the right track and given the well-established finding that ToM is impaired in autistic populations (e.g. Baron-Cohen, 2000), then autistic traits should exert an effect on SI generation (mediated by ToM skill), which was not evident in our results. First, we should mention that the ToM explanation is tentative and speculative and that conclusive support for this explanation requires further empirical and conceptual work. Moreover, we have two points to make concerning the possibility raised by the reviewer. To start with, the questionnaire employed in the current study (AQ) does not directly test ToM. Even though it includes some items that might be interpreted as tapping ToM (e.g. *I find it easy to work out what someone is thinking or feeling just by looking at their face*, or *I find it difficult to work out people*'*s intentions*), there are only a few of these items (three out of the 50 items are about understanding feelings and intentions, one is about being polite and another one about understanding jokes). Hence, the questionnaire measures ToM only indirectly and unreliably. Given this indirectness and unreliability in measuring ToM, one should not expect a strong correlation between AQ scores (either overall or in specific scales) and rate of SI responses, even if the above-mentioned pattern of relations between autistic traits, ToM, and SI interpretation were true. To the contrary, a relation between ToM and rate of SI generation might be observed only if a direct and reliable measure of ToM is employed (such as, for example, the perspective-taking tasks used in Apperly et al. (2010), Keysar et al. (2000), or Keysar et al. (2003)). The second point we wish to make is more general. Given that our sample included neuro-typical adults who do not have clinically significant levels of autism, it is possible that variability in ToM skill in this population does not stem from degree of autistic traits but from other reasons. We believe this is an important point that requires some elaboration. The inference from high AQ scores to low ToM scores is based on the assumption that individuals without an autism diagnosis who score high in AQ have substantial similarities with individuals clinically diagnosed with autism (who also score high in AQ and who have been reported to have low ToM scores). However, this is just an assumption, because high AQ scores need not be causally related to ToM. For instance, low ToM and high AQ scores may both be downstream effects of the same underlying impairment, i.e. clinically-diagnosed autism but causally unrelated. Indeed, this point has been made in terms of visual perception and cognition in a sample of individuals without an autism diagnosis but with high AQ scores and a sample of individuals with autism diagnosis and high AQ scores (Gregory and Plaisted-Grant, 2013). It was reported that the non-diagnosed group did not show the visual search profile that is consistently associated with individuals diagnosed with autism. The upshot of these observations is that findings from clinically-diagnosed populations should be treated with caution and not automatically generalised to the typical population based on scores from screening questionnaires.

online cognitive measure (N400 modulation). While it is possible that individual differences in N400 modulation (indicating that only some participants compute SIs online) reflect patterns of participant responses in behavioural tasks such as ours (with some participants consistently rejecting and some consistently accepting the critical items), it is just as possible to maintain that this is not the case (since this was not tested empirically). Thus, this explanation cannot be excluded and is open to further investigation.

Finally, we should note that we do not consider the cognitive-based and personality-based accounts of SI generation to be mutually exclusive. The finding that WM was a significant predictor of SI generation in the current experimental context does not preclude the possibility that, in different communicative contexts, personality factors might also affect SI interpretation. The task used in our experiments was a verification task. Although verifying sentences is part of what people do in everyday communication (see Bott et al., 2012), it is possible that the absence of a more realistic communicative setting discouraged participants from projecting their own personality characteristics onto the speaker and from using their own attitudes and personal expectations to shape their interpretation of the critical utterances. The study by Feeney and Bonnefon (2013) suggests, precisely, that personality characteristics might also come into play in the process of interpreting scalar terms, given richer and more naturalistic story contexts in which participants are invited to take the protagonists' perspective and in which interpersonal relations between characters make considerations of politeness and honesty highly relevant.

Acknowledgments

Parts of this research have been funded by an ESRC Experimental Pragmatics Network in the UK grant (XPrag-UK; RES-810-21-0069) to all authors and an Alexander Onassis Foundation scholarship for graduate studies to the first author.

Appendix A. Sample pictorial stimuli from the under-informative statement task

Image 1: Initial image

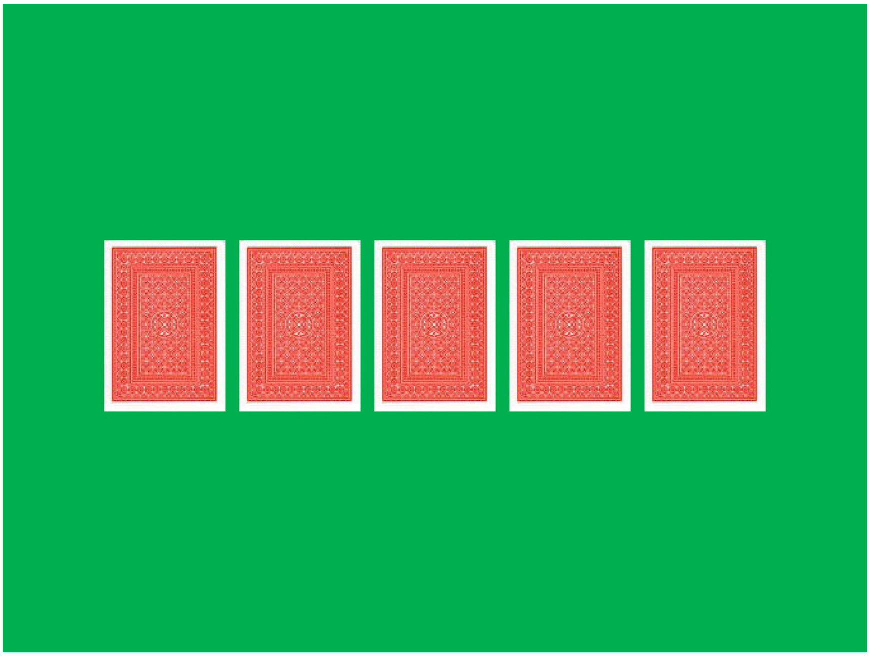


Image 2: Revealed display, 3/5 stars situation

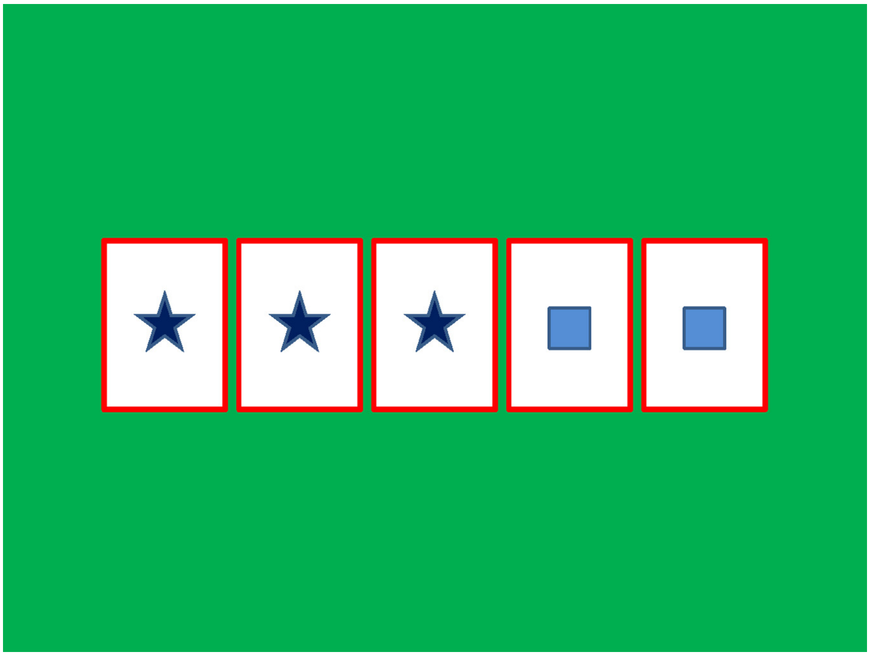


Image 3: Revealed display, 5/5 stars situation

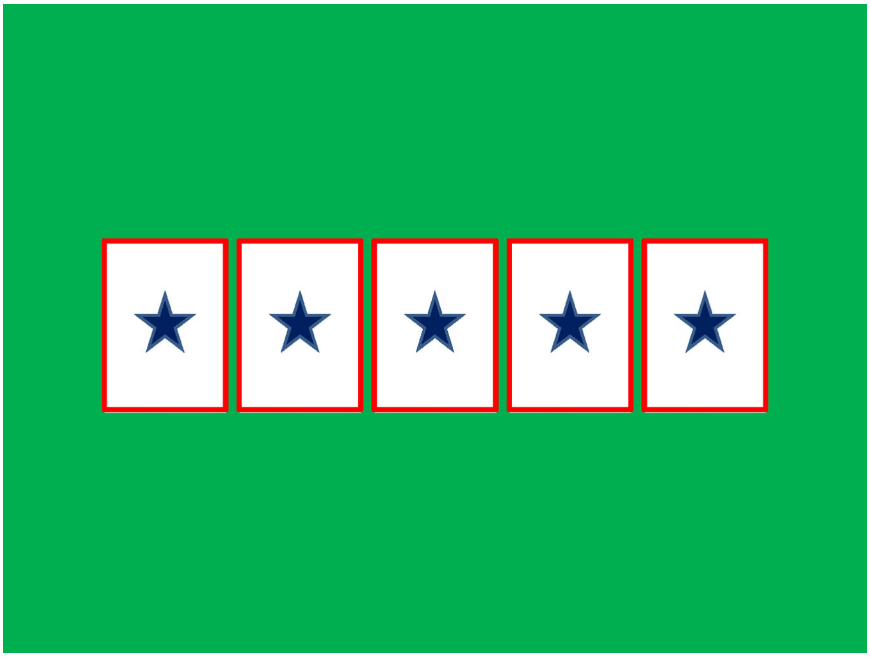
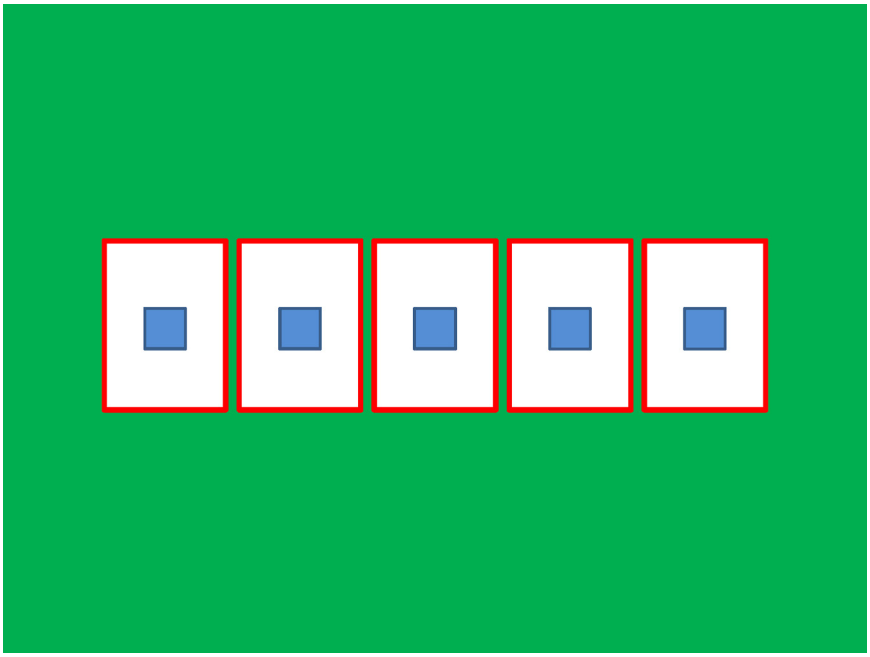


Image 4: Revealed display, 0/5 stars situation



Appendix B. List of linguistic stimuli from the under-informative statement task

Target utterance	Visual display	Utterance status	Expected response
There are hearts on some of the cards.	5/5	True but under-informative	True/False
There are hearts on some of the cards.	3/5	True and informative	True
There are hearts on some of the cards.	0/5	False	False
There are hearts on all of the cards.	5/5	True and informative	True
There are hearts on all of the cards.	3/5	False	False
There are hearts on none of the cards.	3/5	False	False
There are hearts on none of the cards.	0/5	True and informative	True
There are suns on some of the cards.	5/5	True but under-informative	True/False
There are suns on some of the cards.	3/5	True and informative	True
There are suns on some of the cards.	0/5	False	False
There are suns on all of the cards.	5/5	True and informative	True
There are suns on all of the cards.	3/5	False	False
There are suns on none of the cards.	3/5	False	False
There are suns on none of the cards.	0/5	True and informative	True
There are stars on some of the cards.	5/5	True but under-informative	True/False
There are stars on some of the cards.	3/5	True and informative	True
There are stars on some of the cards.	0/5	False	False
There are stars on all of the cards.	5/5	True and informative	True
There are stars on all of the cards.	3/5	False	False
There are stars on none of the cards.	3/5	False	False
There are stars on none of the cards.	0/5	True and informative	True
There are squares on some of the cards.	5/5	True but under-informative	True/False
There are squares on some of the cards.	3/5	True and informative	True
There are squares on some of the cards.	0/5	False	False
There are squares on all of the cards.	5/5	True and informative	True
There are squares on all of the cards.	3/5	False	False
There are squares on none of the cards.	3/5	False	False
There are squares on none of the cards.	0/5	True and informative	True
There are rings on some of the cards.	5/5	True but under-informative	True/False
There are rings on some of the cards.	3/5	True and informative	True

Appendix B (*Continued*)

Target utterance	Visual display	Utterance status	Expected response
There are rings on some of the cards.	0/5	False	False
There are rings on all of the cards.	5/5	True and informative	True
There are rings on all of the cards.	3/5	False	False
There are rings on none of the cards.	3/5	False	False
There are rings on none of the cards.	0/5	True and informative	True
There are moons on some of the cards.	5/5	True but under-informative	True/False
There are moons on some of the cards.	3/5	True and informative	True
There are moons on some of the cards.	0/5	False	False
There are moons on all of the cards.	5/5	True and informative	True
There are moons on all of the cards.	3/5	False	False
There are moons on none of the cards.	3/5	False	False
There are moons on none of the cards.	0/5	True and informative	True

Appendix C. List of experimental materials from the Sentence repetition task

Instructions

In each trial of this experiment you are going to hear a pre-recorded voice saying a sentence. Listen carefully, and when the voice stops, you will have to repeat the sentence exactly as said. There is a practice item before the main part of the experiment. Now press the down arrow key to continue.

Sentences used in the Sentence repetition task

The children got in line for lunch (practice item).
Some children have libraries in their school.
As the sun was setting we put up our tents.
After she ate her sandwich, Martha drank her milk.
Because her right hand was in a cast, she had difficulty writing.
Each morning the birds sing in the trees outside my window.
The woman standing by the man in the green jacket is my aunt.
Long lines of people were waiting at the entrance to the theatre.
The teenagers on our block are raising money to help build a community centre.
Because a storm was approaching, we packed the picnic food away in the big basket.
The fresh vegetables were sliced and arranged in the bowl to make a crisp salad.
Next Wednesday, at 2 o clock in the afternoon, our baseball team will play in a tournament at the stadium.
Mary's books on socialism and Marx were left on the shelf that Jake had failed to attach safely to the wall.
The politician denied any involvement in the scandal, accusing his enemies in the media of making up stories.
Bill had difficulty following the recipe, which called for the sugar to be heated to a very precise temperature.
This coming week John will attend the seminar by the history professor whose books he had admired since he was at university.
Over the course of human history, many people have entered the European peninsula through maritime routes in search of food, shelter and a brighter future.

Note: This is a modified version of the NEPSY sentence repetition battery (Korkman et al., 1999).

References

Apperly, I.A., Riggs, K.J., Simpson, A., Chiavarino, C., Samson, D., 2006. Is belief reasoning automatic? Psychol. Sci. 17 (10), 841–844.

Apperly, I.A., Carroll, D.J., Samson, D., Humphreys, G.W., Qureshi, A., Moffitt, G., 2010. Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. Q. J. Exp. Psychol. 63 (6), 1201–1217.

Baddeley, A.D., 2000. The episodic buffer: a new component of working memory? Trends Cogn. Sci. 4 (11), 417–422.

Baddeley, A.D., Hitch, G., 1974. Working memory. In: Bower, G. (Ed.), The Psychology of Learning and Motivation. Academic Press, New York, pp. 47–90.

Baron-Cohen, S., 2000. Theory of mind and autism: a fifteen year review. In: Baron-Cohen, S., Tager-Flusberg, H., Cohen, D.J. (Eds.), Understanding Other Minds: Perspectives from Developmental Cognitive Neuroscience. Oxford University Press, Oxford, pp. 3–20.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E., 2001. J. Autism Dev. Disord. 31, 5–17.

Bergen, L., Grodner, D.J., 2012. Speaker knowledge influences the comprehension of pragmatic inferences. J. Exp. Psychol.: Learn. Mem. Cogn. 38 (5), 1450–1460.

Bernstein, D.M., Thornton, W.L., Sommerville, J.A., 2011. Theory of mind through the ages: older and middle-aged adults exhibit more errors than do younger adults on a continuous false belief task. Exp. Aging Res. 37 (5), 481–502.

Bezuidenhout, A., Cutting, J.C., 2002. Literal meaning, minimal propositions, and pragmatic processing. J. Pragmat. 34 (4), 433–456.

Bott, L., Noveck, I.A., 2004. Some utterances are underinformative: the onset and time course of scalar inferences. J. Mem. Lang. 51 (3), 437–457.

Bott, L., Bailey, T.M., Grodner, D., 2012. Distinguishing speed from accuracy in scalar implicatures. J. Mem. Lang. 66 (1), 123–142.
Breheny, R., Katsos, N., Williams, J., 2006. Are generalised scalar implicatures generated by default? An on-line investigation into the role of context in generating pragmatic inferences. Cognition 100 (3), 434–463.
Breheny, R., Ferguson, H.J., Katsos, N., 2013. Taking the epistemic step: toward a model of on-line access to conversational implicatures. Cognition 126 (3), 423–440.
Cacciari, C., Tabossi, P., 1988. The comprehension of idioms. J. Mem. Lang. 27, 668–683.
Carston, R., 1998. Negation, 'presupposition', and the semantics/pragmatics distinction. J. Linguist. 34, 309–350.
Chiappe, D.L., Chiappe, P., 2007. The role of working memory in metaphor production and comprehension. J. Mem. Lang. 56 (2), 172–188.
Chierchia, G., 2004. Scalar implicatures, polarity phenomena, and the syntax/pragmatics interface. In: Belletti, A. (Ed.), Structures and Beyond. Oxford University Press, pp. 39–103.
Chierchia, G., Crain, S., Guasti, M.T., Gualmini, A., Meroni, L., 2001. The acquisition of disjunction: evidence for a grammatical view of scalar implicatures. In: Do, A.H.J., Domingues, L., Johansen, A. (Eds.), Proceedings of the 25th Boston University Conference on Language Development. Cascadilla Press, Somerville, MA, pp. 157–168.
Chierchia, G., Fox, D., Spector, B., 2011. The grammatical view of scalar implicatures and the relationship between semantics and pragmatics. In: Maienborn, C., von Heusinger, K., Portner, P. (Eds.), Semantics: An International Handbook of Natural Language Meaning. Berlin, Mouton de Gruyter.
Conway, A., Kane, M., Butting, M., Hambrick, D.Z., Wilhelm, O., Engle, R., 2005. Working memory span tasks: a methodological review and user's guide. Psychon. Bull. Rev. 12 (5), 769–786.
De Neys, W., Schaeken, W., 2007. When people are more logical under cognitive load: dual task impact on scalar implicature. Exp. Psychol. 54, 128–133.
Dieussaert, K., Verkerk, S., Gillard, E., Schaeken, W., 2011. Some effort for some: further evidence that scalar implicatures are effortful. Q. J. Exp. Psychol. 64 (12), 2352–2367.
Feeney, A., Bonnefon, J.F., 2013. Politeness and honesty contribute additively to the interpretation of scalar expressions. J. Lang. Soc. Psychol. 32 (2), 181–190.
Fox, D., 2007. Free choice and the theory of scalar implicatures. In: Sauerland, Uli, Stateva, P. (Eds.), Presupposition and Implicature in Compositional Semantics. Palgrave Macmillan, New York, pp. 537–586.
Geurts, B., 2010. Quantity Implicatures. Cambridge University Press, Cambridge.
Giora, R., 2003. On Our Mind: Salience, Context and Figurative Language. Oxford University Press, Oxford.
Goldberg, L., Johnson, R.J., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R., Gough, H.G., 2006. The international personality item pool and the future of public-domain personality measures. J. Res. Personal. 40, 84–96.
Gregory, B.L., Plaisted-Grant, K.C., 2013. The autism-spectrum quotient and visual search: shallow and deep autistic endophenotypes. J. Autism Dev. Disord. 1–10.
Grice, H.P., 1989. Studies in the Way of Words. Harvard University Press, Cambridge, MA.
Grodner, D., Sedivy, J., 2005. The effect of speaker-specific information on pragmatic inferences. In: Pearlmutter, N., Gibson, E. (Eds.), The Processing and Acquisition of Reference. MIT Press, Cambridge, MA.
Grodner, D., Klein, N., Carbary, K., Tanenhaus, M., 2010. ''Some'' and possibly all, scalar inferences are not delayed: Evidence for immediate pragmatic enrichment. Cognition 116, 42–55.
Guasti, M.T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., Meroni, L., 2005. Why children and adults sometimes (but not always) compute implicatures. Lang. Cogn. Process. 20 (5), 667–696.
Henry, J.D., Phillips, L.H., Ruffman, T., Bailey, P.E., 2012. A meta-analytic review of age differences in theory of mind. Psychol. Aging. http://dx.doi.org/10.1037/a0030677
Hill, E.L., 2004. Executive dysfunction in autism. Trends Cogn. Sci. 8 (1), 26–32.
Hirschberg, J., 1991. A Theory of Scalar Implicature. Garland, New York.
Horn, L., 1972. On the Semantic Properties of Logical Operators in English (PhD thesis). University of California, Los Angeles.
Horn, L., 2004. Implicature. In: Horn, L.R., Ward, G. (Eds.), The Handbook of Pragmatics. Blackwell Publishers, Malden, MA, pp. 3–28.
Horn, L., 2005. The border wars: a neo-Gricean perspective. In: Turner, K., von Heusinger, K. (Eds.), Where Semantics Meets Pragmatics. Elsevier.
Huang, Y., Snedeker, J., 2009. On-line interpretation of scalar quantifiers: insight into the semantics-pragmatics interface. Cogn. Psychol. 58, 376–415.
John, O.P., Naumann, L.P., Soto, C.J., 2008. Paradigm shift to the integrative big-five trait taxonomy: history, measurement, and conceptual issues. In: John, O.P., Robins, R.W., Pervin, L.A. (Eds.), Handbook of Personality: Theory and Research. Guilford Press, New York, pp. 114–158.
Kane, M.J., Hambrick, D.Z., Tuholski, S.W., Wilhelm, O., Payne, T.W., Engle, R.W., 2004. The generality of working memory capacity: a latent-variable approach to verbal and visuo-spatial memory span and reasoning. J. Exp. Psychol. Gen. 133, 189–217.
Katsos, N., Bishop, D.V., 2011. Pragmatic tolerance: implications for the acquisition of informativeness and implicature. Cognition 120 (1), 67–81.
Katsos, N., Smith, N., 2010. Pragmatic tolerance or a speaker–comprehender asymmetry in the acquisition of informativeness. In: Franich, K., Iserman, K.M., Keil, L.L. (Eds.), Proceedings of the 34th Annual Boston Conference in Language Development. Cascadilla Press, MA, USA, pp. 221–232.
Keysar, B., Barr, D.J., Balin, J.A., Brauner, J.S., 2000. Taking perspective in conversation: the role of mutual knowledge in comprehension. Psychol. Sci. 11, 32–38.
Keysar, B., Lin, S., Barr, D.J., 2003. Limits on theory of mind use in adults. Cognition 89, 25–41.
Korkman, M., Kirk, U., Kemp, S., 1999. NEPSY: A Developmental Neuropsychological Assessment. The Psychological Corporation, San Antonio, TX.
Levinson, S., 2000. Presumptive Meanings. MIT Press, Cambridge, MA.
Lin, S., Keysar, B., Epley, N., 2010. Reflexively mindblind: using theory of mind to interpret behavior requires effortful attention. J. Exp. Soc. Psychol. 46 (3), 551–556.

Marty, P., Chemla, E., 2013. Scalar implicatures: working memory and a comparison with ''only''. Front. Psychol. 4.
Marty, P., Chemla, E., Spector, B., 2013. Interpreting numerals and scalar items under memory load. Lingua 133, 152–163.
Mashal, N., 2013. The role of working memory in the comprehension of unfamiliar and familiar metaphors. Lang. Cogn. 5 (4), 409–436.
McElree, B., Nordlie, J., 1999. Literal and figurative interpretations are computed in parallel. Psychon. Bull. Rev. 6, 486–494.
Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., Wager, T.D., 2000. The unity and diversity of executive functions and their contributions to complex ''Frontal Lobe'' tasks: a latent variable analysis. Cogn. Psychol. 41 (1), 49–100. http://dx.doi.org/10.1006/cogp.1999.0734
Nieuwland, M.S., Ditman, T., Kuperberg, G.R., 2010. On the incrementality of pragmatic processing: an ERP investigation of informativeness and pragmatic abilities. J. Mem. Lang. 63 (3), 324–346.
Noveck, I.A., Posada, A., 2003. Characterising the time course of an implicature. Brain Lang. 85, 203–210.
Novick, J.M., Hussey, E., Teubner-Rhodes, S., Harbison, J.I., Bunting, M.F., 2013. Clearing the garden-path: improving sentence processing through cognitive control training. Lang. Cogn. Neurosci. 29 (2), 186–217. http://dx.doi.org/10.1080/01690965.2012.758297
Papafragou, A., Musolino, J., 2003. Scalar implicatures: experiments at the semantics/pragmatics interface. Cognition 86, 253–282.
Poirier, M., Martin, J.S., Gaigg, S.B., Bowler, D.M., 2011. Short-term memory in autism spectrum disorder. J. Abnorm. Psychol. 120 (1), 247.
Pouscoulous, N., Noveck, I., Politzer, G., Bastide, A., 2007. A developmental investigation of processing costs in implicature production. Lang. Acquis. 14, 347–376.
Richmond, L.L., Thorpe, M., Berryhill, M.E., Klugman, J., Olson, I.R., 2013. Individual differences in autistic trait load in the general population predict visual working memory performance. Q. J. Exp. Psychol. 66 (6), 1182–1195.
Schneider, D., Lam, R., Bayliss, A.P., Dux, P.E., 2012. Cognitive load disrupts implicit theory-of-mind processing. Psychol. Sci. 23 (8), 842–847.
Schuh, J.M., Eigsti, I.-M., 2012. Working memory, language skills, and autism symptomatology. Behav. Sci. 2 (4), 207–218.
Shapiro, A.M., Murphy, G.L., 1993. Can you answer a question for me? Models of processing indirect speech acts. J. Mem. Lang. 32, 211–229.
Simon, J.R., 1969. Reactions towards the source of stimulation. J. Exp. Psychol. 81, 174–176.
Sperber, D., Wilson, D., 1986/1995. Relevance: Communication and Cognition. Blackwell, Oxford.
Stroop, J.R., 1935. Studies of interference in serial verbal reactions. J. Exp. Psychol. 18, 643–662.
Tomlinson Jr., J.M., Bailey, T.M., Bott, L., 2013. Possibly all of that and then some: scalar implicatures are understood in two steps. J. Mem. Lang. 69 (1), 18–35. http://dx.doi.org/10.1016/j.jml.2013.02.003
Unsworth, N., Redick, T.S., Heitz, R.P., Broadway, J.M., Engle, R.W., 2009. Complex working memory span tasks and higher-order cognition: a latent-variable analysis of the relationship between processing and storage. Memory 17 (6), 635–654.
Wechsler, D., 1944. The Measurement of Adult Intelligence, 3rd ed. The Williams and Wilkins Company, Baltimore.
Wechsler, D., 1999. Wechsler Abbreviated Scale of Intelligence (WASI). San Antonio Pearson Assessment, Inc..

Dr Kyriakos Antoniou has a PhD in Theoretical and Applied Linguistics from the University of Cambridge and is currently a Postdoctoral Researcher at the Université libre de Bruxelles in Belgium. His main research interests are in experimental pragmatics, bilingualism, and its influence on children's cognitive and pragmatic skills.

Dr Chris Cummins is a Chancellor's Fellow in the Department of Linguistics and English Language at the University of Edinburgh. Prior to this he worked at Bielefeld University as part of SFB 673 ''Alignment in Communication''. His recent publications include a monograph, Constraints on Numerical Expressions (OUP, 2015). Other research interests include presupposition, metaphor, and the interpretation of indirect speech acts.

Dr Napoleon Katsos is a Senior Lecturer at the Department of Theoretical and Applied Linguistics of the University of Cambridge. He is interested in theories of lexical and sentential meaning informed by empirical data from acquisition and processing studies in monolingual and bilingual children and adults.