Quaderni di Linguistica e Studi Orientali QULSO
Working papers in Linguistics and Oriental Studies

Firenze University Press
https://oajournals.fupress.net/index.php/bsfm-qulso

Citation: G. Starr (2025) Maintaining Context or Reacting to It? Cognitive Control Strategy and Awareness of the "Question Under Discussion" in Scalar Implicature Derivation. *Qulso* 11: pp. 101-126. doi: http://dx.doi.org/10.13128/QULSO-2421-7220-18664

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Competing Interests: The Author(s) declare(s) no conflict of interest.

Maintaining Context or Reacting to It? Cognitive Control Strategy and Awareness of the "Question Under Discussion" in Scalar Implicature Derivation*

Glenn Starr
University of Wisconsin, Milwaukee (<starrg@uwm.edu>)

Abstract:

This experimental study examines whether the Question Under Discussion (QUD) modulates the comprehension of pragmatic implicatures in context. Using brief two-sentence communicative exchanges, participants are asked to judge the acceptability of a target *some* response whose felicity is influenced by *all* and *any* quantifiers in the preceding QUD as well as an accompanying picture graphic. This work also investigates the extent to which context maintenance ability affects native English speakers' awareness of the contextual cues in the target QUD conditions. Context maintenance ability was indexed using a Dot Pattern Expectancy (DPX) task, allowing examination of its potential predictive role in participants' judgments. Data collected from 39 native English speakers reveal significant differences in rating behavior in the two target conditions. Participants rated *some* responses as less acceptable following *all*-QUDs than *any*-QUDs, indicating sensitivity to contextual cues that prime scalar implicatures. Interestingly, older adults who showed slower and less accurate performance on context-dependent trials in the DPX task were more likely to distinguish between QUD conditions in their judgements. These findings suggest that certain individuals rely on a more reflective, reactive processing strategy when reasoning about scalar inference generation. Rather than proactively maintaining contextual information throughout, they engage cognitive control to incorporate context cues only after ambiguity or interference arises.

Keywords: Scalar Implicature, QUD Sensitivity, Context Awareness, Context Maintenance Ability, Cognitive Control, DPX Task

* I extend very special thanks to Drs. Jacee Cho, Thomas Purnell, and Eric Raimy for help with conceptualization and experimental design. A debt of gratitude is also extended to Kathy Starr, Peggy Cerrito, and Linda Wilford for their support in participant recruitment as well as Dr. Jonathan Jibson for help in recording the audio dialogues. I am also very grateful to the two reviewers who were so helpful with their comments and suggestions which helped greatly to improve the final version of the paper.

1. Introduction

Scalar inferences represent a linguistic property that is of growing interest to researchers in the fields of semantics, pragmatics, and psycholinguistics (Bott & Noveck 2004; Noveck & Sperber 2012; Noveck 2018). The most frequently studied scalar inference involves the *some* phrase from the < *some, all* > scale. The quantifier *some* is inherently ambiguous in that it contains two distinct meanings. It can be interpreted in its strictly semantic or logical form to mean *at least one and possibly all* or as a common pragmatic inference meaning *some but not all* (Horn 1972; Chierchia et al. 2019). The pragmatically enriched meaning implicates the negation of other higher order terms such as *most* or *all*. Consider the following:

(1) Some politicians are corrupt.
(2) Not all politicians are corrupt.
(3) All politicians are corrupt.

In most scenarios, people typically infer that (1) means (2) even though *some* here can also be interpreted to mean *some and possibly all*. The meaning generated in (2) is a type of conversational implicature that emerges as a result of an assumption that is inferentially derived by the hearer (Grice 1989; Horn 2006; Levinson 2000). If the speaker meant to convey the meaning in (3), they would have used the quantifier *all* in order to be maximally informative. Deriving additional meaning in conversations beyond the literal meaning of words necessitates adherence to mutually understood conversational maxims that regulate rational discourse (Grice 1989). One of these, the Quantity maxim, holds that speakers need to make their utterances as informative as is required for a given situation. Thus, when a listener hears (1), they will consider whether the speaker meant for the informationally stronger terms *most* or *all* to apply. Since these terms were not used, and assuming the speaker is being rational and forthright, the *some but not all* inference will be generated.

Psycholinguists are curious about how scalar inferences are derived and what cognitive mechanisms contribute to their generation (a.o., Huang & Snedeker 2009; Katsos & Cummins 2010). Since human communication involves more than just the straight encoding and decoding of words or messages in their strictly semantic form, pragmatic inference processing involves the ability to interpret the meaning of an utterance beyond its literal interpretation. This inevitably requires a degree of both pragmatic and contextual awareness on the part of listeners in communicative exchanges.

How scalar implicatures (SIs) are generated, and whether they are computed by default or through other more cognitively demanding means, is a subject of interest to researchers in semantics and pragmatics. Neo-Gricean researchers such as Levinson (2000) state that SI derivation is automatic in most scenarios. They argue that a default pragmatically enriched interpretation of *some* as *some but not all* evolved out of the demands for efficiency in communication. This camp holds that only in very rare cases does *some* carry a literal meaning and arriving at this interpretation requires careful evaluation of both speaker intent and context in order to cancel the default reading. However, this reversal will come with a cost.[1]

[1] See also Chierchia (2004, 2006) and Chierchia et al. (2019) who propose that SI computation involves a covert focus operator **O**, assigned by the grammar, capable of taking scope over any constituent with a propositional meaning. In example (1), this theory suggests that the silent grammatical operator **O** takes scope over the sentence containing the scalar term *some* resulting in the negation of the alternative proposition with *all* and thus computing the SI.

Relevance theorists hold an opposing view stating that implicature generation is costly in terms of processing resources because meaning for the *some* phrase must first be extracted from a context before the derivation process can proceed. This necessitates a narrowing down of factors present in each communicative scenario in order to exclude atypical readings of *some*. The resulting interpretations are further guided by expectations of relevance (Sperber & Wilson 1995; Wilson & Sperber 2006; Noveck & Sperber 2012). Relevance theorists disagree with the Levinsonian view of default SI generation and there are psycholinguistic studies which offer evidence supporting the perspective that implicature generation is effortful (Bott & Noveck 2004; De Neys & Schaeken 2007; Marty, Chemla, & Spector 2013, among others).[2,3]

A common thread across the different theories is the emphasis on attentiveness to conversational elements that shape the interpretation of *some*. According to Pickering and Garrod's (2004) interactive alignment model, successful communication relies on interlocutors' mutual coordination and shared situational awareness of communicative goals. This alignment is facilitated by a shared, often implicit, understanding of linguistic structures and discourse expectations. Through recursive alignment at multiple linguistic levels, speakers and listeners co-construct meaning. The present study explores participants' sensitivity to semantic-pragmatic cues that influence the interpretation of *some*, as well as the cognitive traits underlying this sensitivity. Effective interpretation in such contexts requires maintaining contextual awareness as the communicative exchange unfolds.

2. Context and QUD in Implicature Generation

According to Roberts (2006, 2012), "Questions under discussion" (QUDs) represent pertinent indicators of speaker intent during dialogue. They initiate goal-seeking behavior among participants in conversation and highlight the importance of offering relevant contributions as dialogue unfolds. Goal-oriented discourse thus benefits from strategies that aid in the noticing of relevant stimuli, and which facilitate the end-goals of an exchange. These endeavors consist of both sequential and recursive moves since a hearer may have to refer back to previous utterances in order to maintain the contextual grounds upon which a relevant answer can be supplied.

With respect to SIs, the QUD can influence how a subsequent *some* response is interpreted and whether an implicature is warranted.[4] These cues may come in the form of quantifiers such as *none*, *all*, or *any*. Consider the following examples in dialogue:

(4) Speaker A: Are <u>all</u> the students passing the class?
Speaker B: Some students are passing.
(5) Speaker A: Is there <u>any</u> evidence against them?
Speaker B: Some of their identity documents are forgeries.
(adapted from Levinson 2000)

[2] See, however, Ronderos and Noveck (2023) who demonstrate that delays in implicature generation can at least partly be explained by the need for participants to adjust to new speakers and to think about the speaker's informative intentions with respect to underinformative phrases.

[3] Most prior work on SIs has focused on contrasting Neo-Gricean and Relevance Theory accounts. More recently, constraint-based approaches (e.g., Degen & Tanenhaus 2015) have emerged, proposing that implicature derivation reflects an interactive, probabilistic process shaped by multiple graded cues.

[4] See Cummins (2017) for a theoretical review and account of how QUD influences the availability of quantity implicatures involving *some* as well as numerical expressions.

The quantifiers *any* and *all* in (4-5) influence expectations of implicature relevance. In (4), the response from Speaker B here expressly institutes a *not all* implicature as relevant due to the *all* quantifier in the QUD. The *all* cue makes explicit a whole and unpartitioned set of students. Thus, in the subsequent '*Some students are passing*' response, it can be inferred that *some* refers to the specific subset *some but not all* of the students. The *all* priming component is absent in (5), however. The quantifier *any* does not have a place on the < *some*, *most*, *all* > scale.

Although the *any* QUD in (5) may not prime an implicature reading in the same way that *all* can, it does not mean that the hearer necessarily refrains from deriving the inference. Nothing in the context directly inhibits it. Nevertheless, Levinson (2000) writes that the inference will likely be suspended. The reason being that "intuitively, A is only interested in whether there is at least some evidence against the criminals; given A's question, all that is relevant is the possession of at least some evidence" (p. 51). Levinson (2000) holds that any implicatures generated under these circumstances would thus be considered under the auspices of Grice's maxim of Relevance as a Relevance implicature, rather than of Quantity. The *all* QUD, however, because of its proximity in the context to its scalemate *some*, suggests consideration of scalar competitors and thus applies selection pressure towards generation of the *some but not all* inference. Referring to Grice's (1989) Quantity maxim, since Speaker B in (4) could have chosen to use *all* in the response but chose not to, the implicature is more expressly licensed in this context.

2.1 "All" versus "any" QUD and the role of executive function in SI generation

Several studies in recent years have demonstrated how cues in the experimental design affect the interpretation and processing of *some* phrases. Adapting the methodology of Breheny et al. (2006), Politzer-Ahles and Fiorentino (2013) carried out a self-paced reading study which measured the processing costs associated with comprehending *some* in context. The only difference in the two target conditions rest with the use of *all* or *any.* In (6), the context biases readers towards an "upper-bound" (e.g., *some but not all*) reading for *some*. As explained, what is relevant is whether *all* is true in relation to *some* while in (7), the "lower-bound" condition, what is relevant is whether *any* is true. As such, the *some of* phrase in (7) is less likely to be interpreted as *some but not all*.

(6) Upper-bound: Mary was preparing to throw a party for John's relatives. She asked whether *all* of them were staying in his apartment. John said that <u>**some of them**</u> were. He added that <u>**the rest**</u> would be staying in a hotel.

(7) Lower-bound: Mary was preparing to throw a party for John's relatives. She asked whether *any* of them were staying in his apartment. John said that <u>**some of them**</u> were. He added that <u>**the rest**</u> would be staying in a hotel.

Politzer-Ahles and Fiorentino (2013) found that no extra processing cost was incurred at the *some of* phrase in either condition. However, in (7), participants recorded significant slow-downs at *the rest* segment as compared to the same segment in (6). It was argued that *all* in (6) expressly licenses an upper-bound interpretation for *some* so by the time participants parsed the underlined segments, the inference had already been derived. In (7), an upper-bound reading is made irrelevant and the existence of a subset of relatives only comes to mind at *the rest* section. These results confirmed their prediction that the use of *any* in a preceding context is less likely to prime *some* as having an upper-bound *some but not all* interpretation.

Another experimental study examining the role of more explicit QUD in SI derivation comes from Degen's (2013) dissertation. In experiment 2a, Degen sought to test for the possibility that QUD modulates implicature generation rates. Participants read a picture story and were told to imagine themselves in a store full of experimental gumball machines that were programmed to verbalize the amount of gumballs they released after each use. Two target QUD conditions were created. The first included *all* in the question (e.g., *Did I get all of the gumballs?*) and the other used *none* (e.g., *Did I get none of the gumballs?*). After presentation of the question, participants were then asked to agree or disagree to the machine's *You got some of the gumballs* utterance. The accompanying picture showed that all gumballs had been released. In the *all* condition, participants supplied pragmatic responses 92% of the time, meaning they disagreed with the machine's statement. This rate dropped to 50% in the *none* QUD condition, confirming Degen's prediction that QUD could be used to manipulate the likelihood of implicature generation.

Using a related sentence-picture verification task, Ronai and Xiang (2021) presented participants with short question-and-answer dialogues about a collection of shapes that were accompanied by visual displays. In the target condition, all shapes were blue. Only the responder was said to see the shapes, while the questioner asked either *Are any of the shapes blue?* or *Are all the shapes blue?* In response, the target reply always used a *some* statement (e.g., *Some of the shapes are blue*). Participants judged each response as either "Good" or "Not Good." Consistent with Degen (2013), judgements varied systematically by QUD type: in the *all*-QUD condition, 75% of responses were judged "Not Good," compared to just 50% in the *any*-QUD condition.

Starr & Cho (2022) investigated native and second language (L2) speakers' sensitivity to *all* and *any* cues in explicit questions using a related design. In target trials, two interlocutors were depicted having a conversation about a set of five objects, all of which were of the same shape and color. The questioner, who could not see the shapes, asked the question (e.g., *Are all/any of the squares red?*) to which the hearer responded with a *some* phrase (e.g., *Some shapes are red*). Using a 7-point Likert scale, participants were instructed to judge the acceptability of the response after considering the question and visual in each scenario. Native speakers differentiated significantly in rating behavior, rating *any* QUD contexts higher than *all* ones. The Mandarin Chinese L2-English learners, however, did not, suggesting that native and L2 speakers may attend to different information types during language processing.

Yang, Minai, and Fiorentino (2018) employed a story-sentence matching task also investigating the role of *all* and *any* QUD in SI derivation. In target trials, two interlocutors were depicted having a conversation about a set of four objects that had all been manipulated in the same way. The QUD included either an *all* or *any* quantifier (e.g., *Have you cut all/any of the steaks?*) whereupon the responder replied with an underinformative *some* response (e.g., *I cut some steaks*). Participants were asked to provide a judgement about the appropriateness of the response using a Likert scale.

Yang and colleagues also included several cognitive and personality assessments in the study. While psycholinguistic research suggests that working memory as well as age are factors that contribute to differing interpretations of underinformative *some* phrases in isolation (De Neys and Schaeken 2007; Dieussaert et al. 2011; Marty and Chemla 2013; Marty et al. 2013; Antoniou et al. 2016), Yang and colleagues explored how executive function, reasoning ability, language skills, and personality traits influence awareness of *some* in context. Participants completed tasks measuring working memory, context maintenance ability, and attentional control. The researchers then compiled these values into a "cognitive resources" score. A further personality assessment gathered information about pragmatic reasoning ability. Their results revealed that both metrics affected the likelihood of a participant incorporating the *any* and *all* cues in their judgements. That is, those with higher executive function and pragmatic reasoning ability scores registered greater rating differences in *all* vs. *any* QUD conditions.

Though intriguing, Yang et al.'s (2018) scoring paradigm presents two key concerns. First, by merging scores from three distinct executive function tasks into a composite "cognitive resources" score, the individual contributions of each skill are obscured. As a result, it becomes challenging to isolate and evaluate the specific impact of context maintenance ability, or any of the other measured aptitudes, on their own with respect to the incorporation and retention of cues essential for SI generation.

While tests of working memory, context maintenance, and attentional control all assess aspects of cognitive control, they differ in the specific cognitive functions they measure. Working memory tests, for example, emphasize temporary storage and manipulation of information over short periods rather than proactive maintenance of task-relevant contextual material or resolution of conflicting information (Baddeley 2012, 2020). Attentional control tasks measure the ability to resolve interference and suppress automatic but incorrect responses (Stroop 1935; Kane & Engle 2003; Scarpina & Tagini 2017). Finally, context maintenance ability measures aptitude in maintaining and sustaining context over a delayed period in order to influence later task-relevant behavior (Servan-Schreiber et al. 1996; Cohen et al. 1999; Boudewyn et al. 2015). The current work focuses on this last skill specifically, investigating whether the ability to maintain and use contextual information correlates with sensitivity to QUD in the interpretation of *some*.

3. Methods

The following experiment employs an acceptability judgement task (AJT) wherein the felicitousness of a *some* response is affected by a preceding QUD containing *all* or *any* along with an accompanying picture graphic. This study also seeks to examine whether context maintenance ability affects awareness of the QUD. Two research questions are addressed. The first asks whether *all* and *any* QUDs modulate the interpretation of *some*. If participants incorporate these context cues, it is predicted that an *all*-QUD will promote SI generation while the *any* counterpart will temper it. Second, this experiment asks whether context maintenance ability affects one's propensity to notice and incorporate signals issued by each QUD. If this ability affords individuals with increased awareness of contextual information, it is expected that they will display increased sensitivity to the different QUD conditions.

3.1 Participants

A total of 39 native English speakers took part in the experiment. This group consisted primarily of university age undergraduate and graduate students studying at a large midwestern university in the United States. A smaller contingent of older working age participants from around the country also elected to participate. The mean age for the group was 35.4 years (*SD* = 20.7), with ages ranging from 18 to 77 years.[5]

[5] Though 26 of the 39 total participants were below age 40 with 25 of them aged 31 or younger, 13 participants were over the age of 50 with 8 aged over 65. Therefore, age effects are analyzed in the statistical analysis.

3.2 Materials

3.2.1 Sentence-Picture Judgement task

Adopting Starr and Cho's (2022) experimental design, participants completed an AJT in Qualtrics consisting of a series of short dialogues about a picture depicting five shapes all of the same kind (e.g., circles, diamonds, squares, or triangles). The exchanges included one question and a response presented in audio form. Prior to beginning the experiment, participants were instructed that the questioner in each exchange was unable to see how many colored shapes there were. Depending on the trial condition, either zero, three, or all five of the shapes were filled in with the same color (e.g., blue, green, orange, purple, red, or yellow). The length of each exchange did not exceed five seconds. The included picture with five shapes was displayed below the audio clip. Once the clip had finished playing, participants could replay the sequence before advancing to a judgement screen that included only the response and the picture. Below them, the question *How acceptable is the response to the question?* appeared above a 7-point Likert scale with "1" meaning "unacceptable" and "7" meaning "acceptable" (Figure 1).

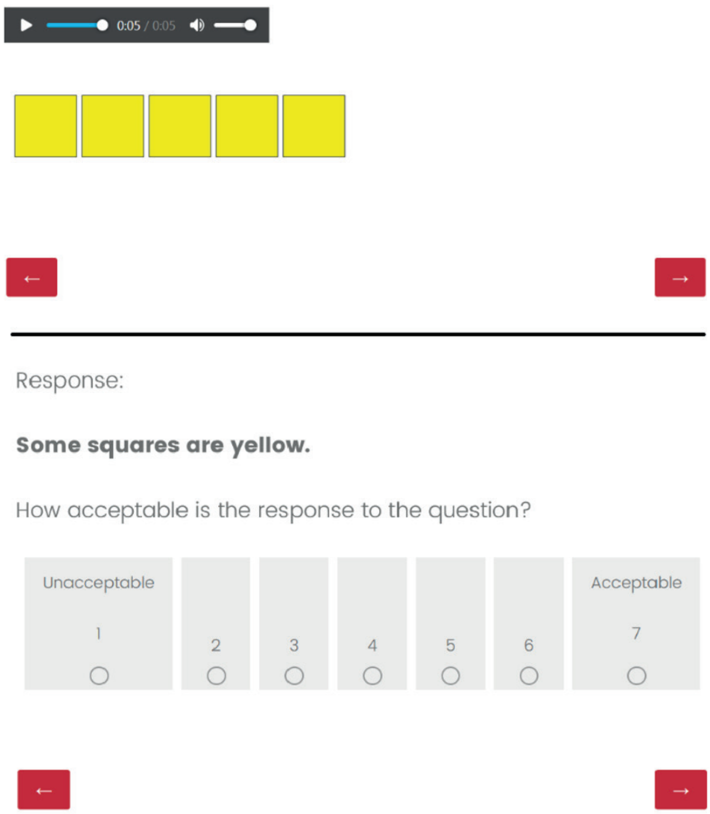

Figure 1. Sample display of a target trial in the judgement task

The experiment used a 3x2 factorial design with two within-subjects factors: QUD and Picture. The QUD factor had two levels: *all*-QUDs (e.g., *Are all the squares yellow?*) and *any*-QUDs (e.g., *Are any of the squares yellow?*). The Picture factor had three levels, each depicting a different proportion of colored shapes (Table 1). Each trial consisted of a question (QUD), followed by a picture, and then a *some* response (e.g., *Some of the squares are yellow*). The *5/5 picture level* is of particular interest because, when paired with an *all*-QUD, it creates a context where a pragmatic interpretation of *some* (i.e., *not all*) becomes relevant. In this condition, the *some* response is predicted to garner lower acceptability ratings. By contrast, when the *5/5 picture level* appears with an *any*-QUD, the context does not strongly prime the inference. This would theoretically increase the likelihood that the *some* response means *some and possibly all* which would result in higher acceptability ratings.

QUD type / Picture type	all	any
5/5 colored shapes (target)	Are all the squares yellow? Some squares are yellow.	Are any of the squares yellow? Some squares are yellow.
3/5 colored shapes (control - true)	Are all the squares yellow? Some squares are yellow.	Are any of the squares yellow? Some squares are yellow.
0/5 colored shapes (control - false)	Are all the squares yellow? Some squares are yellow.	Are any of the squares yellow? Some squares are yellow.

Table 1. Six test conditions in a 3x2 factorial design: "QUD" type (*all* vs. *any*) and "Picture" type (*5, 3,* and *0*)

The control conditions included pictures levels depicting 3 out of 5 or 0 out of 5 colored shapes, each paired with a some response. Filler trials replaced the all and any quantifiers in the QUD with five, three, or none, and the some response was replaced with the corresponding numeral. In filler-true trials, the QUD, response, and picture numerically aligned. In filler-false trials, there was a mismatch in at least one of these components. The control and filler trials served to ensure participant attentiveness. No data were excluded, as all participants completed the task as instructed. Six unique lists were created using a Latin square design. Each list contained four items per target and control condition (eight target and 16 control trials total), along with the same 24 filler trials (12 true, 12 false), yielding 48 trials per list.

3.2.2 Context Maintenance: Dot Pattern Expectancy (DPX) Task

Context maintenance ability was assessed using the Dot Pattern Expectancy (DPX) task, a variant of the AX-Continuous Performance Task (AX-CPT) widely used in cognitive psychology (Servan-Schreiber et al., 1996). Individuals with strong context maintenance skills can retain relevant contextual cues in working memory and apply them to guide interpretation in tasks involving lexical or referential ambiguity (Boudewyn et al., 2015; Cohen et al., 1999). In this framework, contextual information refers to task-relevant cues that must be actively maintained and utilized to support appropriate responses (Cohen et al.,1999: 120). Successful context maintenance therefore relies on both sustained working memory and the ability to evaluate competing inputs to determine which information should be suppressed, updated, or integrated to guide behavior.

The DPX task presents participants with sequential pairs of dot patterns (e.g., a cue followed by a probe) on a computer screen. Participants are instructed to press a designated key only when a specific target pairing appears; for all other pairings (i.e., non-target types), they must press an alternative key. Each trial begins with a cue pattern composed of blue dots centered on a white background. After a brief delay, a probe pattern composed of red dots appears. The task includes four distinct trial types (Table 2). In target AX trials, a target (A) cue is followed by a target (X) probe, prompting the target response. The remaining three trial types require the non-target response. AY trials present a target A cue followed by a non-target Y probe. BX trials involve a non-target B cue followed by a target X probe. Finally, BY trials feature a non-target B cue and a non-target Y probe. These conditions are designed to assess participants' context maintenance ability by requiring participants to distinguish between cues and probes across varying contexts.

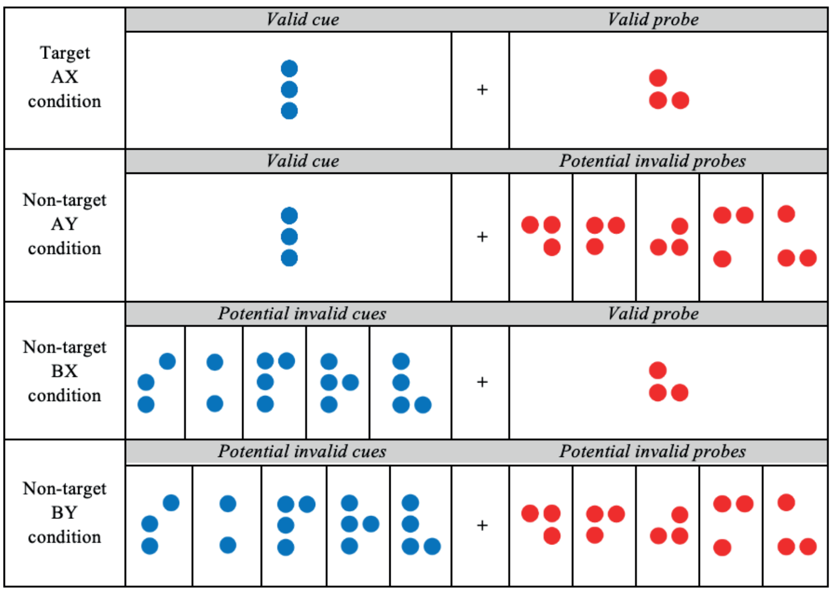

Table 2. Examples of the target cue-probe (AX) condition and all non-target cue-probe (AY, BX, BY) trial pairs present in the experiment

The DPX task was administered using E-Prime (Psychology Software Tools 2021). Prior to the start of the task, participants read instructions and completed 10 practice trials. The main task consisted of 128 trials (88 AX, 16 AY, 16 BX trials, and 8 BY trials) distributed randomly in the presentation lists. In each trial sequence, the cue first appeared on screen for 1000ms, followed by an interstimulus interval of 2000ms. The interval screen displayed a fixation cross centered against the white background. Next, the probe pattern appeared for 500ms, and participants had the opportunity to press the appropriate key during this time or for up to 1500ms after the probe vanished. A feedback screen then appeared for 1500ms notifying participants whether they were correct in their choice. A subsequent intertrial interval screen appeared for 1200ms whereupon the trial sequence began anew.

Since the initial target A cue appears in the majority of trials (81.25%), it provides some initial information to the participant about the likelihood that the following dot pattern will make up a target trial. In the AY condition, the presence of the valid A cue creates an expectancy in the user that a valid X probe will follow. This expectation arises because of the larger proportion of AX trials that appear throughout the task (68.75%). The target A cue in AY trials thus necessitates careful evaluation of the ensuing probe, but participants must also contend with the probabilistic context that a valid X probe is more likely to appear. Thus, the AY sequence adds a degree of difficulty since participants must ultimately reject the pairing after the appearance of the non-target Y probe.

For those particularly sensitive to context, the AY condition produces longer reaction times (RTs) and increased error rates (MacDonald et al. 2005; Jones et al. 2010). Upon encountering a non-target Y probe in AY trials, the context sensitive individual must expend resources to override a habituated target probe response that had built up over time due to the higher probability ratio of target X probes in AX trials. In the AY condition, the processor is contending with unplanned interference which results in increased cognitive load as it maneuvers to incorporate the novel context. For those who possess comparatively poorer context processing ability, however, it is argued that they would register fasters RTs and fewer errors overall since less interference is introduced by the presence of a target A cue (Braver 1997; Braver & Cohen 2001).

In both BX and BY trials, the initial appearance of a non-target cue B signals that the trial does not require a target response. Consequently, these trials typically yield faster reaction times (RTs) and lower error rates than AX or AY trials. However, BX trials are also informative for assessing context maintenance ability since target X probe creates a prepotent tendency in some individuals to respond as if the trial were a target (AX). Accurate performance on BX trials therefore requires participants to maintain the contextual information provided by the B cue to override this prepotent response. Individuals with stronger context maintenance skills are generally quicker to recognize B cues as non-target and are therefore less prone to erroneously conclude that the X probe is part of the target pair. Elevated error rates and slower RTs in BX trials signal reduced sensitivity to contextual information and impaired ability to inhibit misleading cues (Braver et al. 2005). In contrast, BY trials, which lack the prepotent X probe, are less cognitively demanding and tend to elicit near-ceiling accuracy regardless of individual differences in context processing.

3.2.2.2 Proactive and Reactive Modes of Context Processing

At its core, the DPX task is a test of context maintenance. That is, it tests the ability to keep the cue (A or B) in mind over a short delay and use it to interpret the probe (X or Y). This ability is critical for goal-directed behavior, but the DPX can also tell us about the different cognitive control strategies participants use when encountering novel context or unexpected interference.

Braver and colleagues, for example, advance a "dual mechanisms of control" (DMC) theory which posits that there are two distinct modes of cognitive control characterized by *proactive* or *reactive* processing attributes (Braver 2012; Braver et al. 2005; Braver et al. 2009). The proactive control mode is defined by its tendency towards early selection and maintenance of goal-relevant information for later use in cognitively demanding tasks. The reactive control mode is characterized by late correction in which task-relevant information is activated only when interference occurs.

Proactive control refers to the anticipatory maintenance of goal-relevant information, akin to thinking ahead or strategic preparation. In the DPX task, this involves actively maintaining the A cue in working memory to prepare for the expected X probe. Individuals who utilize this control mode display characteristic performance patterns across trial types. Specifically, AX trials typically yield fast RTs and low error rates, as the maintained A cue facilitates a prepared response to X. In AY trials, the A cue creates an expectancy for X, but the appearance of Y violates this prediction, requiring response inhibition. Thus, AY trials often produce slower RTs and increased errors. In contrast, BX and BY trials begin with a non-target B cue, which signals that no target response will be required, rendering the identity of the probe irrelevant. Accordingly, both BX and BY trials generally elicit faster RTs and minimal error rates, as the cues allow early disengagement of the target-driven response preparation.

Reactive cognitive control engages only when conflict or ambiguity arises, rather than sustaining task-relevant context in anticipation of future stimuli. In the DPX task, those who adopt this strategy show evidence of limited cue maintenance and greater reliance on the probe to determine the appropriate response. As a result, even AX trials may yield slower RTs or reduced accuracy due to the absence of proactive anticipation of context. In AY trials, where no expectancy for X is generated, the Y probe is easily identified as non-target, leading to fast RTs and low error rates. BX trials, however, are particularly diagnostic of reactive control. Without maintaining the non-target B cue, participants with reactive profiles may either respond to X as if in a target AX trial or enter a contextual reevaluation period to restore from working memory the representation of the target pair. The ensuing conflict triggers *retroactive* control processes, such as conflict monitoring and attempted cue retrieval, which nevertheless slow responses and also increase the likelihood of errors due to the limited time frame in each trial. In contrast, BY trials involve no expectancy violation or conflict and are typically rejected quickly and accurately (see Table 3).

Trial Type	Proactive Control Mode		Reactive Control Mode	
	RTs	*Errors*	*RTs*	*Errors*
AX	Fast	Low	Moderate	Moderate
AY	Slower	Higher	Fast	Low
BX	Fast	Low	Slower	Higher
BY	Fast	Low	Fast	Low

Table 3. Expected diagnostic signatures of proactive vs. reactive control modes in DPX task performance

3.2.2.3 Context Maintenance Ability Scoring and Metrics

DPX task assessments can be done in different ways. In Yang et al. (2018), a DPX task similar to the one described above was used. Yang and colleagues also gathered scores from a numerical Stroop task (Bush et al. 2006) to assess attentional control ability and working memory aptitude was recorded using a Count Span exercise (Conway et al. 2005). Recall that values from these three metrics were combined into a composite "cognitive resources" score for use as a singular coefficient in mixed effects regression modelling. The value they generated from the DPX task was determined via a *d*-prime calculation method (Swets & Sewall 1963; Cohen et al. 1999; Henderson et al. 2012).

To capture the *d*-prime score, a participant's hit rate (or accuracy in terms of percentage) from the AX condition is recorded and transformed into a *z*-score value. In BX trials, a false alarm rate (or percentage of inaccurate responses) is recorded. The *z*-score transformed false alarm rate from the BX condition is then subtracted from the hit rate in the AX condition to get the resulting *d*-prime score.[6] Those who register higher *d*-prime scores are argued to be more sensitive to context. However, this method ignores error rates in AY and BY trials as well as RTs in AX, AY, BX, and BY trials even though these values can be useful indicators of an individual's context maintenance aptitude and cognitive control profile on their own (Braver 1997; Braver & Cohen 2001; Braver et al. 2005; MacDonald et al. 2005; Jones et al. 2010).

To investigate how context maintenance ability relates to sensitivity to QUD and the interpretation of scalar *some*, the present study examines RTs, error rates, and *d*-prime scores across all four DPX trial conditions. The additional analysis of RTs and error rates allows for investigation of claims in cognitive psychology regarding performance in each condition in relation to general context sensitivity and cognitive control. Proactive individuals, who maintain contextual cues in working memory, should therefore exhibit accurate responses in AX and BX trials and this should be reflected by larger rating differentials in the *all* versus *any* QUD condition. There is also the possibility that reactive control, characterized by slower RTs in BX trials and greater conflict-driven processing, may also support context sensitivity. Thus, we specifically test whether these cognitive control signatures predict differentiation in scalar *some* ratings under *all* and *any* QUDs.

3.3 Procedure

Participants were provided with links to the survey and to the DPX task. The survey welcome screen informed users that they would be provided with some example and practice items before beginning the main task. The example trials contained elements and questions that were unrelated to the linguistic properties under investigation but were designed to familiarize users with the ensuing discourse scenarios and the Likert scale rating system. Participants were instructed to complete the task using either headphones or speakers and that they should pay close attention to the question and the response as well as the picture on the screen when considering their judgements. When completed, participants were instructed to start the DPX task.

[6] Hit rates of 100% in the AX condition are generally adjusted to 1-1/160 and false alarm rates of 0% are corrected to 1/16 (Cohen et al. 1999). This adjustment was applied to applicable cases for for the present work as well.

4 Results

4.1 Data Cleaning and Analysis

Prior to beginning statistical analysis, the data were trimmed and organized in the following ways. Rather than choosing an overall short/long cutoff method for the RTs in the DPX task, trimming of the RT data involved removing all recorded values from each participant that were more than two standard deviations above or below their mean. Using a universal cutoff method with an overall short or long RT value limit was not applied since this method would inevitably result in slower (or faster) subjects having meaningful data points excised from their datasets. The ± two standard deviation cutoff approach allows for outliers to be safely removed relative to the base reaction time for each subject (Ratcliff 1993). This procedure resulted in the loss of 2.6% of the DPX data.

Inferential statistical analysis was conducted by compiling several Cumulative Link Mixed Models (CLMMs) using the *ordinal* package (Christensen 2023). CLMMs are ideal for regression modelling with Likert scale responses because they respect the ordinal nature of the data. By using a probit link function, CLMMs can model the log-transformed odds of one rating exceeding another, such as a score of 5 versus 4. Additionally, they incorporate random effects to account for variations at both the subject and item levels.

Prior to model fitting, all control and filler trials were removed from the dataset to isolate the effects of the QUD, RT, and error rate variables on the target condition. All subsequent models set the Likert scale rating as the dependent variable. To avoid accumulation of too many predictors in any one model, a series of CLMMs were created testing for interaction effects between the QUD, age, and individual DPX task metrics on response behavior.[7] Thus, the primary model included QUD, a *z*-score transformed *d*-prime score, age, and their interactions as fixed effect predictors. Ancillary models included QUD, a *z*-score transformed mean RT or error rate value from each condition in the DPX task, age, and their interactions as fixed effect covariates. All models included random intercepts for "Subject" and "Item". The QUD factor, consisting of the two levels *all* and *any*, was treatment coded with *all* set as the reference level. Each individual model was thus specified as follows:

$$\text{Response} \sim \text{QUD} * \text{Age} * \textit{DPX metric} \text{ (e.g., } d\text{-prime, mean RT, or error rate)} + (1|\text{Subject}) + (1|\text{Item})$$

4.2 Descriptive Statistics

It was predicted that target conditions containing an *all*-QUD, a picture with five of five of the same colored shapes, and an accompanying underinformative *some* response phrase would yield lower ratings than the target *any*-QUD condition. Table 4 provides a preliminary view of response behavior. On average, participants provided a meaning rating of 2.73 (*SD* = 1.56) in the *all* target condition and a higher mean rating of 3.45 (*SD* = 1.76) for *any* trials. Ratings in

[7] Due to significant variation in age, an age covariate was included as well. Its addition increased the r-squared values and lowered the values of the Akaike Information Criterion (AIC) in all models in which it was added. We return to the significance of this variable in the discussion.

both control categories demonstrate that participants adhered to the task as instructed. In the control-true condition showing only three of five of the same colored shapes, a *some* response is appropriate and so both *all* and *any* QUD conditions received similarly high ratings, 6.31 and 6.79, respectively. In the control-false condition, the accompanying picture shows zero of five colored shapes so the subsequent *some* response is incorrect. Ratings here hovered between 1.06 and 1.08. The violin plot in Figure 2 shows the distribution of every rating given in target trials and the median value for each condition. "2" was the middle value for all trials while "3" was the middle value for any trials in the dataset. Ratings in the *all* condition tend to accumulate towards the bottom of the scale but ratings in *any* trials are more widely distributed.

	QUD	Picture	Mean Rating	SD
Participants (n = 39)	all	5/5	2.73	1.56
	any	5/5	3.45	1.76
	all	3/5	6.31	1.40
	any	3/5	6.79	0.51
	all	0/5	1.06	0.35
	any	0/5	1.08	0.35

Table 4. Mean rating and standard deviations to the "QUD" (all/any) and "Picture" (5/5, 3/5, 0/5) conditions (*Note*: The target category is in gray)

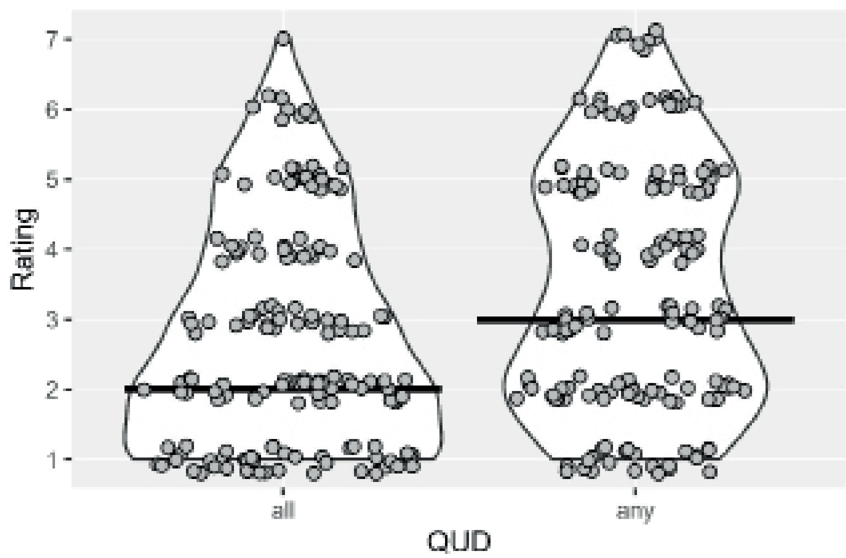

Figure 2. Distribution of ratings in target conditions of the AJT task (*Note*: The black bars represent median values)

4.3 Inferential Analyses

4.3.1 D-prime and QUD Sensitivity

Analyzing these results statistically, the primary CLMM model tested the fixed effects of QUD, *d*-prime, age, and their interactions on response behavior.[8] Table 5 summarizes the model output. A robust effect of QUD emerged (β = 2.74, *SE* = 0.38, *z* = 7.30, *p* < .001), indicating that ratings were significantly higher in the *any*-QUD condition compared to *all*-QUD trials. Age also significantly predicted rating behavior, with older participants assigning lower ratings in the *all*-QUD condition (β = -2.89, *SE* = 0.74, *z* = -3.93, *p* < .001). Additionally, higher *d*-prime scores were associated with increased ratings in the *all*-QUD condition (β = 2.53, *SE* = 1.10, *z* = 2.29, *p* = .022).

Several interactions qualified these simple main effects. A significant QUD x Age interaction (β = 2.91, *SE* = 0.40, *z* = 7.23, *p* < .001) showed that ratings in the *any*-QUD condition increased with age, yielding a crossover pattern in which older participants gave lower ratings in *all* trials but higher ratings in *any* trials. The QUD x *d*-prime interaction was also significant (β = -2.13, *SE* = 0.62, *z* = -3.46, *p* < .001), indicating that participants with higher *d*-prime scores gave lower ratings in *any*-QUD trials, effectively *reducing* the contrast between QUD conditions. This suggests that *d*-prime did not predict greater sensitivity to QUD context, as it failed to produce the expected interaction pattern.

A significant Age x *d*-prime interaction (β = 3.80, *SE* = 1.43, *z* = 2.67, *p* = .008) suggested that among older participants, higher *d*-prime scores were associated with increasing ratings in the *all*-QUD condition. Finally, a significant three-way interaction between QUD, Age, and *d*-prime (β = -3.09, *SE* = 0.82, *z* = -3.77, *p* < .001) showed that the reduction in QUD-based differentiation was most pronounced among older participants with higher *d*-prime scores. Post hoc pairwise comparisons with Bonferroni correction confirmed these effects.

Fixed Effects	*Estimate*	*SE*	*z*	*p*-value
QUD	2.739	0.375	7.301	<0.001***
Age at testing	-2.894	0.737	-3.929	<0.001***
d-prime	2.528	1.102	2.293	0.022*
QUD x Age at testing	2.912	0.403	7.231	<0.001***
QUD x *d*-prime	-2.134	0.618	-3.455	<0.001***
Age x *d*-prime	3.804	1.425	2.669	0.008**
QUD x Age at testing x *d*-prime	-3.090	0.818	-3.777	<0.001***
Random Effects	*Variance*	*SD*		
Subject (*n* = 39)	12.645	3.556		
Item (*n* = 24)	0.252	0.502		

Marginal R^2=0.24; Conditional R^2=0.85
Note: Number of observations (*n* = 318) – *p < 0.05, ** p < 0.01, *** p < 0.001

Table 5. Fixed effects estimates from CLMM with QUD, *d*-prime, and age as predictors of AJT ratings

[8] Visit the OSF link (<https://osf.io/z28ah/?view_only=26e7ef8f78d546c2b5119947ef8e6a92>) for test items and R-scripts.

The *d*-prime model revealed that fixed effects (QUD, Age, d-prime, and their interactions) accounted for 23.8% of the variance in acceptability ratings (marginal $R^2 = 0.238$), while the full model including subject- and item-level variability accounted for 84.5% of the variance (conditional $R^2 = 0.845$). These values suggest substantial individual differences in rating behavior, above and beyond the effects of experimental predictors.

4.3.2 QUD Sensitivity as Indexed by Error Rates and Reaction Times

Much like the main model, all RT and error rate models revealed a consistent interaction between QUD and age, such that increased age was associated with greater sensitivity to the QUD. In the AX and BX RT models, significant three-way interactions for QUD x AX RT x Age ($\beta = 0.79$, $SE = 0.33$, $z = 2.40$, $p = .016$) and QUD x BX RT x Age ($\beta = 0.69$, $SE = 0.32$, $z = 2.19$, $p = .028$) indicated that older adults who responded more slowly on these trials exhibited greater differentiation in rating behavior between QUD conditions. This pattern suggests that increased deliberation during contextually routine AX but also in interference-laden BX trials may reflect activation of enhanced or controlled processing in older adults. In contrast, AY and BY trials yielded no reliable RT effects. Marginal R^2 values for all RT models ranged from 0.15 to 0.24 while conditional R^2 values remained steady at 0.83, again suggesting significant subject-level variance.

No significant interactions emerged between QUD, error rate, and age in the AX, AY, or BY error rate models. However, the BX model revealed two noteworthy interactions indicating that error patterns and age jointly predict QUD sensitivity in older adults (Table 6). A significant negative interaction between BX error rate and age ($\beta = -8.25$, $SE = 3.79$, $z = -2.18$, $p = .029$) showed that higher error rates among older individuals were associated with lower ratings in *all*-QUD trials. Additionally, a significant three-way interaction (QUD x BX error rate x Age: $\beta = 7.57$, $SE = 3.61$, $z = 2.10$, $p = .036$) indicated that older adults with elevated BX error rates exhibited greater rating differentiation between QUD types. This suggests that increased susceptibility to interference may promote greater context sensitivity in older participants, possibly reflecting compensatory or strategic processing. Post hoc pairwise comparisons with Bonferroni correction supported these effects. The BX model's marginal R^2 was 0.52, with a conditional R^2 of 0.91. Corresponding R^2 values for the AX, AY, and BY models ranged from 0.15-0.46 (marginal) and 0.82-0.89 (conditional).

Fixed Effects	*Estimate*	*SE*	*z*	*p*-value
QUD	3.858	1.105	3.491	<0.001***
BX error rate	-5.986	2.842	-2.106	0.035*
Age at testing	-4.189	1.541	-2.718	0.007**
QUD x BX error rate	5.398	2.698	2.001	0.045*
QUD x Age at testing	4.242	1.410	3.008	0.003**
BX error rate x Age at testing	-8.251	3.789	-2.178	0.029*
QUD x Age at testing x *d*-prime	7.574	3.614	2.096	0.036*
Random Effects	*Variance*	*SD*		
Subject (*n* = 39)	14.124	3.758		
Item (*n* = 24)	0.227	0.476		

Marginal R^2=0.52; Conditional R^2=0.91
***Note*: Number of observations (*n* = 318) – *p < 0.05, ** p < 0.01, *** p < 0.001**

Table 6. Fixed effects estimates from CLMM with QUD, BX error rate, and age as predictors of AJT ratings

To visualize the significant interactions between context sensitivity, cognitive control, and age in predicting sensitivity to QUD manipulation, an interaction plot was generated (Figure 3). The visualization displays predicted QUD sensitivity at different age levels across a range of *d*-prime, BX error rate, and BX RT values. QUD sensitivity was scored via a *z*-score differential value, created by taking a participant's mean *z*-score rating in all trials and subtracting that from the mean rating in any trials.[9] Positive values indicate higher ratings overall in *any* vs. *all* QUD conditions. Zero values show no differentiation between conditions and negative ratings demonstrate higher ratings for all compared to any trials. The plot depicts the predicted *z*-score differential at three levels of Age: the mean (average age 35.4 years), −1 SD (younger participants), and +1 SD (older participants). This allows for visual inspection of how the relationship between DPX task performance and QUD sensitivity varied with participant age, offering insight into age-related shifts in cognitive control dynamics.

[9] *Z*-score differentials were used to categorize participants instead of raw mean rating differences to account for individual variability in Likert scale use. *Z*-scoring standardizes individual response patterns and enhances comparability across participants by controlling for differences in response distributions (see Sprouse & Almeida 2013).

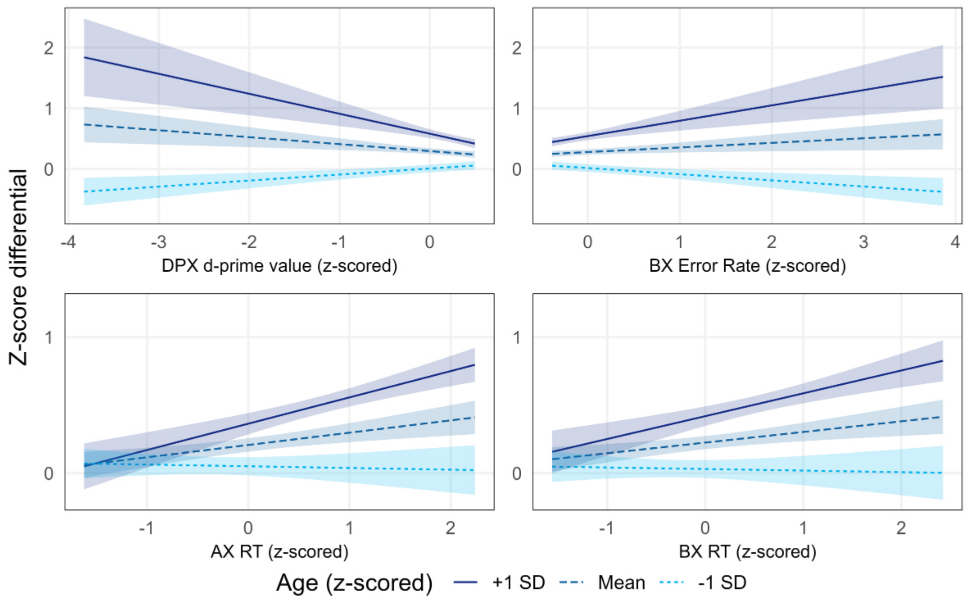

Figure 3. QUD Sensitivity as Indexed by *d*-prime, BX Error Rate, AX and BX Reaction Time Ranges

In older adults, poor *d*-prime scores are associated with a larger *z*-score differential value and as *d*-prime scores improve, this value decreases. In younger adults, this trend reverses: their poorer *d*-prime scores yield negative *z*-score differential values, but these show an increase into positive territory as *d*-prime rises. We see the opposite dynamic with BX error rates. With increasing age, the *z*-score differential increases along with more errors in the BX condition. However, younger people register a downward sloping relationship under these conditions. Additionally, increases in AX and BX RTs result in greater differentiation between *all* and *any* QUD ratings for the older group. This suggests that prolonged periods of contextual reevaluation reflect greater QUD sensitivity with increasing age. In younger individuals, increases in AX and BX RTs do not yield such effects as the *z*-score differential value remains static across all RT ranges.

5. Discussion

This study adopted Starr and Cho's (2022) acceptability judgement task whereby *all* and *any* quantifiers were embedded within QUDs asking about 5/5 objects sharing the same characteristics. With an *all*-QUD, the interpretation of the *some* response is rendered underinformative and was predicted to yield more "unnatural" ratings in comparison to an *any*-QUD where *some* could be interpreted literally to mean *some and possibly all*. Using a similar design, Yang and colleagues (2018) found that personality factors and cognitive traits influenced response behavior. However, executive function in their study was tabulated via a composite "cognitive resources" score which combined results from attentional control, context maintenance ability, and working memory tasks. The current work adopted this approach but focused specifically on the role of context maintenance ability to isolate its effect with respect to contextual awareness in SI derivation. Using a DPX task, a *d*-prime score was used in the analysis together with information from response time and error rates across all trial types.

Two questions motivated the present work, each of which were provoked by findings in psycholinguistics and cognitive psychology research. The first concerned the effect of the QUD on the comprehension of pragmatic implicatures involving *some*. Several novel experiments in recent years have returned notable findings regarding the effect of referential context and quantifier cues on implicature generation (Breheny et al. 2006; Degen 2013; Politzer-Ahles & Fiorentino 2013; Politzer-Ahles & Husband 2018; Yang et al. 2018; Ronai & Xiang 2021; Starr & Cho 2022). In line with previous studies, participants demonstrated sensitivity to *all* and *any* QUD in their interpretation of *some*. Likert scale ratings with *all*-QUD were lower in comparison to ratings in the *any* condition since *all* is a competing scalar term for *some* in the provided context. This priming effect is absent with *any* QUDs, however, because *any* and *some* do not share a scalar or degree relation. Participants were sensitive to this implication and rated the *any* condition as more acceptable.

The second research question concerned the extent to which context maintenance ability contributed to awareness of the provided cues. Successful conversation necessitates incorporation of relevant contextual information regarding the intentions of each interlocutor. Moreover, this information must be continually updated and tracked as the communicative exchange evolves (Pickering & Garrod 2004; Roberts 2006, 2012). All current theories on SI recognize the role that context plays in generating a scalar inference. Apart from Yang et al. (2018), previous studies involving the processing and interpretation of *some* phrases did not assess the degree to which aptitude in context maintenance specifically might modulate contextual awareness. This study's results suggest specific aspects of context maintenance, and more particularly cognitive control, predict sensitivity to QUD, but also that age modulates these processing patterns.

5.1 Age and QUD Awareness

A significant effect of "age" was found during statistical analysis, showing an association between older adults and increasing rating differences between the QUD conditions. That is, older participants were more likely to rate *all* trials lower and *any* trials higher, demonstrating heightened awareness of how cues in the preceding question influence the *some* response. This was an unexpected result given research in experimental pragmatics suggesting that age-related declines in executive function may affect older adults' pragmatic abilities (Bambini et al. 2021; Mazzaggio et al. 2023). It is further understood, for example, that diminished cognitive capacity affects comprehension of non-literal utterances such as irony (Cordonier et al. 2020) and retrieval of linguistic knowledge (Craik & Bialystok 2006).

Research in experimental pragmatics demonstrates that older adults have trouble in understanding figurative utterances in the form of humor (Bischetti et al. 2023) and metaphor (Mashal et al. 2011). They also experience difficulty in processing and understanding statements of implied meaning such as sarcasm (Phillips et al. 2015) and irony (Mazzaggio et al. 2023). Phillips et al. (2015) for example, found that older adults had more difficulty understanding videos containing sarcastic exchanges than middle-aged or young adults. This, even though all three groups performed similarly in the comprehension of stories where sincere interactions were depicted. These studies indicate that advanced age is associated with increased difficulties in incorporating contextual cues in the understanding of implied meaning.

Although Mazzaggio and colleagues (2023) did not assess irony comprehension in older adults, they examined the processing time of ironic statements compared to literal ones using a self-paced reading task. Their results showed that, while both younger (ages 19-27) and older (ages

65-76) adults took longer to process ironic statements than literal ones, older adults exhibited significantly greater processing delays. Mazzaggio et al.'s (2023) study is pertinent to the results of the current work because, by not testing understanding of irony directly and instead focusing only on processing time, their assumption was that the comprehension of implied meaning is still intact in old age. By extension, this suggests older adults may simply require additional processing effort to comprehend, integrate, and process contextual cues during pragmatic inferencing. Connections between these assertions and our results are explored in detail below.

5.2 QUD Sensitivity, Age, and Individual Differences in Cognitive Control Strategy

Though age discrepancy in response behavior was unexpected, approaching these results through a "dual mechanisms of control" (DMC) perspective can help explain the variation. Studies in DMC theory using AX-CPT tasks (the progenitor of the DPX variant) show that there are significant individual differences as well as age-related changes in the updating and maintenance of context information (Braver et al. 2001; Braver et al. 2005; Braver et al. 2009; Schmitt et al. 2014). This variation is primarily expressed via *proactive* versus *reactive* cognitive control modes. In younger adults, context processing is associated with proactive cognitive control where processing behavior is more *anticipatory* in nature and thus more receptive to the early A or B cues which are helpful in designating cue-probe pairs as target or non-target. This processing profile is well-suited for "early preparation of upcoming cognitive tasks that ensures the maintenance of relevant task information to bias response selection" (Schmitt et al. 2014: 202).

In contrast, context processing in older age groups has generally been linked to a reactive cognitive control mode, meaning that they activate cognitive control mechanisms only *after* interference is detected (Braver 2012; Braver et al. 2001; Braver et al. 2005). Recall that reactive control in a DPX task is characterized by limited cue maintenance and more focus on the X probe to determine appropriate responses. Though this strategy helps conserve cognitive resources, it results in slower response times in both AX and BX trials because retroactive control processes must be activated to retrieve cue information. Evidence of this is particularly noticeable in BX trials, where error rates and response slowing represent primary signatures in older people (Braver et al. 2001). These two contrasting processing profiles help explain our results.

In the context of this study, strong context maintenance ability (as indexed by high *d*-prime values) supports a tendency towards proactive processing. This control mode translates into keeping the *all* and *any* QUD cues in mind, applying them early, and rendering a cue-appropriate interpretation of the scalar *some* phrase. Although younger age showed a positive trending association between *d*-prime and QUD sensitivity, the effect was relatively modest (Figure 3). Even at high levels of context maintenance ability, the *z*-score differential values remained close to zero, suggesting that proactive control alone may not be sufficient for robust QUD-based inferencing. This finding may have also been the result of low sample size, however, an issue addressed in the limitations section below.

Conversely, a reactive control mode would be evidenced by late activation of the QUD cue (i.e., only when the underinformative *some* phrase is encountered). That is, when the reactive participant views the "*Some* squares are red" response describing five of five colored shapes, the meaning of *some* is still ambiguous. This creates conflict and interferes with meaning generation as *some* could be interpreted as lower (*some and possibly all*) or upper-bound (*some but not all*). It is only after the *all* or *any* cue from the QUD is retrieved that interference can be alleviated and a judgement rendered. This type of processing strategy is analogous to reactive control behavior in DPX tasks where minimal reliance on A or B cues results in retroactive control processes characterized by conflict monitoring and attempted cue retrieval when the target X probe is encountered. However, this strategy increases RTs and error rates in BX trials, both of which were shown to be significant predictors of QUD sensitivity in this study. Older adults in particular showed significantly *greater* differentiation in rating behavior between QUD conditions when their BX trial error rates and RTs were high. This pattern suggests they may rely on reactive rather than proactive cognitive control during interpretation. Instead of applying control in advance to guide QUD-sensitive processing, older adults appear to recruit it only in response to conflict (i.e., ambiguity). Once engaged, this reactive control mode facilitates retrieval and integration of contextual cues.

The reliance on reactive control may help explain why older adults with higher context maintenance ability (as indexed by high *d*-prime scores) did not show greater sensitivity to the QUD. In fact, as Figure 3 reveals, higher *d*-prime scores in older participants were associated with *reduced* QUD sensitivity and more uniform ratings across *all* and *any* conditions. This suggests that while context maintenance capacity remains relatively intact in later life (Braver et al. 2005), it may not always be flexibly deployed to support complex language processing. More specifically, it may not be leveraged to support flexible, QUD-sensitive interpretation of scalar *some* phrases.

Indeed, older adults often prioritize reliability and efficiency over flexibility in cognitive tasks (Hasher & Zacks 1988; Umanath & March 2014). Thus, even when older individuals exhibit strong *d*-prime performance, they may use this capacity to reinforce a default or well-rehearsed strategy, especially under uncertainty. In the context of this task, such a strategy could involve adopting a stable, context-insensitive interpretation of *some*, regardless of QUD conditions, leading to lower *z*-score differential values. These findings highlight that the timing of control deployment, rather that its availability alone, plays a crucial role in supporting sensitivity to QUD in instances of pragmatic inferencing.

Lastly, it is also important to note that BX error rates and RTs were significant predictors of sensitivity to the QUD in and of themselves, even without the age interaction. This leaves open the possibility that reactive cognitive control may predict QUD and contextual cue sensitivity in instances of scalar implicature computation, irrespective of age. For example, even younger individuals with a reactive processing profile may exhibit heightened context awareness when generating SIs. In this way, the interplay between reactive cognitive control and sensitivity to pragmatic context can very much be considered an individual differences phenomenon. This represents an intriguing avenue for future research.

5.3 Limitations

There are a few issues of concern worth noting about this experiment and its methodology. The experimental design was modeled on that of Starr and Cho (2022) whereby the felicitousness of a *some* response in relation to a picture graphic is argued to be influenced by the preceding *all* or *any* QUD. However, the selection of an appropriate response is not necessarily straightforward. To illustrate, one reason why participants may have assigned low ratings to target trials displaying 5/5 colored shapes is because a *some* response is inappropriate irrespective of the preceding QUD context. For example, if the term *some* is interpreted pragmatically in *all*-QUD trials as *some but not all* then the accompanying picture would simply be factually incorrect (i.e., the response doesn't match the picture). Conversely, if, in the same condition, *some* is interpreted literally to mean *some and possibly all*, the response may not fully address the question. As a result, determining whether the preceding *all*-QUD had any impact becomes challenging. Similarly, high ratings in the *any*-QUD scenario do not necessarily imply that participants interpreted *some* in its logical form. Consequently, while it is plausible that the *all* and *any* QUD influenced the observed differential ratings, it is important to note that the observed response behavior may not be exclusively indicative of SI derivation.

Future research should aim to construct experimental contexts where it is easier for participants to interpret *some* responses and generate clearer interpretive contrasts across QUDs. One suggestion for this would be to use a forced-choice interpretation task that leaves out the picture graphic but asks participants to select whether the lower or upper-bounded meaning of *some* is more likely given the *all* or *any* QUD and response. This would allow researchers to more directly trace the influence of the QUD on interpretation.

Another limitation of the present study is the small sample size ($n = 39$) which may have limited the statistical power to detect more subtle effects and reduced the generalizability of the findings. While the observed patterns offer meaningful insights into age-related differences in QUD sensitivity and cognitive control in relation to the interpretation of scalar *some*, these results should nevertheless be interpreted with caution. Future research should aim to replicate these findings with a larger participant pool to ensure greater reliability and applicability across populations. Additionally, to examine how context maintenance ability manifests differently in different age cohorts, age should be carefully controlled for in future studies.

6. Conclusion

The goal of this article was to explore the possibility that context maintenance ability modulates sensitivity to *all* and *any* contextual cues which can affect the interpretation of underinformative scalar *some* phrases. Using a Question Under Discussion (QUD) paradigm, participants read questions containing either *all* or *any* quantifiers (e.g., “Are all the shapes red?” vs. “Are any of the shapes red?”), followed by a statement containing *some* (e.g., “Some of the shapes are red”) and a visual display showing five out of five red shapes. Scalar implicature derivation was assessed using an acceptability judgement task (AJT), where participants rated these statements on a Likert scale. Crucially, ratings revealed that participants interpreted *some* as *not all* more often in the *all* vs. *any*-QUD trials, demonstrating context-sensitive interpretation. To explore individual differences in context processing, a DPX task was administered post-AJT. From this, *d*-prime scores, error rates, and reaction times were calculated across trial types (AX, AY, BX, BY), each indexing different facets of cognitive control and context sensitivity. Interestingly, participants, especially older adults, with slower RTs in AX and BX trials and higher error rates in BX trials (all signatures of a reactive cognitive control mode) showed greater sensitivity to QUD manipulation in the AJT. This suggests that QUD-sensitive interpretation of *some* may be supported not by proactive cue maintenance, but by reactive cognitive control mechanisms that engage during moments of ambiguity or reanalysis.

References

Antoniou, Kyriakos, Chris Cummins, and Napoleon Katsos. 2016. “Why Only Some Adults Reject Under-Informative Utterances.” *Journal of Pragmatics* 99 (July): 78-95. doi:10.1016/j.pragma.2016.05.001.

Baddeley, Alan. 2012. “Working Memory: Theories, Models, and Controversies.” *Annual Review of Psychology* 63: 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, Alan. 2020. Working memory. In *Memory* (3rd ed.). Routledge.

Bambini, Valentina, Lotte Van Looy, Kevin Demiddele, and Walter Schaeken. 2021. “What Is the Contribution of Executive Functions to Communicative-Pragmatic Skills? Insights from Aging and Different Types of Pragmatic Inference." *Cognitive Processing* 22 (3): 435-52. doi:10.1007/s10339-021-01021-w.

Bischetti, Luca, Irene Ceccato, Serena Lecce, Elena Cavallini, and Valentina Bambini. 2023. "Pragmatics and Theory of Mind in Older Adults' Humor Comprehension." *Current Psychology* 42 (19): 16191-207. doi:10.1007/s12144-019-00295-w.

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 (3): 437-57. doi:10.1016/j.jml.2004.05.006.

Boudewyn, Megan A., Debra L. Long, Matthew J. Traxler, Tyler A. Lesh, Shruti Dave, George R. Mangun, Cameron S. Carter, and Tamara Y. Swaab. 2015. "Sensitivity to Referential Ambiguity in Discourse: The Role of Attention, Working Memory, and Verbal Ability." *Journal of Cognitive Neuroscience* 27 (12): 2309-23. doi:10.1162/jocn_a_00837.

Braver, Todd Samuel. 1997. "Mechanisms of Cognitive Control: A Neurocomputational Model." Ph.D., United States -- Pennsylvania: Carnegie Mellon University. https://www.proquest.com/docview/304369220/abstract/93B7F8BD77584D46PQ/1.

Braver, Todd S., and Jonathan D. Cohen. 2001. "Working memory, cognitive control, and the prefontal cortex: Computational and empirical studies." *Cognitive Processing* 2 (1): 25-55.

Braver, T.S., D.M. Barch, B.A. Keys, C.S. Carter, J.D. Cohen, J.A. Kaye, J.S. Janowsky, et al. 2001. "Context Processing in Older Adults: Evidence for a Theory Relating Cognitive Control to Neurobiology in Healthy Aging." *Journal of Experimental Psychology. General* 130 (4): 746-63.

Braver, Todd S., Ajay B. Satpute, Beth K. Rush, Caroline A. Racine, and Deanna M. Barch. 2005. "Context Processing and Context Maintenance in Healthy Aging and Early Stage Dementia of the Alzheimer's Type." *Psychology and Aging* 20 (1): 33-46. doi:10.1037/0882-7974.20.1.33.

Braver, Todd S., Jessica L. Paxton, Hannah S. Locke, and Deanna M. Barch. 2009. "Flexible Neural Mechanisms of Cognitive Control within Human Prefrontal Cortex." *Proceedings of the National Academy of Sciences* 106 (18): 7351-56. doi:10.1073/pnas.0808187106.

Braver, Todd S. 2012. "The Variable Nature of Cognitive Control: A Dual Mechanisms Framework." *Trends in Cognitive Sciences* 16 (2): 106-13. doi:10.1016/j.tics.2011.12.010.

Breheny, Richard, Napoleon Katsos, and John Williams. 2006. "Are Generalised Scalar Implicatures Generated by Default? An on-Line Investigation into the Role of Context in Generating Pragmatic Inferences." *Cognition* 100 (3): 434-63. doi:10.1016/j.cognition.2005.07.003.

Bush, George, Paul J. Whalen, Lisa M. Shin, and Scott L. Rauch. 2006. "The Counting Stroop: A Cognitive Interference Task." *Nature Protocols* 1 (1): 230-33. doi:10.1038/nprot.2006.35.

Chierchia, Gennaro. 2004. "Scalar Implicatures, Polarity Phenomena, and the Syntax/Pragmatics Interface." In *Structures and Beyond: The Cartography of Syntactic Structures, Volume 3*, edited by Adriana Belletti, 39-103. Oxford University Press. doi:10.1093/oso/9780195171976.003.0003.

Chierchia, Gennaro. 2006. "Broaden Your Views: Implicatures of Domain Widening and the 'Logicality' of Language." *Linguistic Inquiry* 37 (4): 535–90. doi:10.1162/ling.2006.37.4.535.

Chierchia, Gennaro, Danny Fox, and Benjamin Spector. 2019. "10. Scalar Implicature as a Grammatical Phenomenon." In *10. Scalar Implicature as a Grammatical Phenomenon*, 325-67. De Gruyter Mouton. https://www.degruyter.com/document/doi/10.1515/9783110589849-010/html.

Christensen, Rune Haubo Bojesen. 2023. *Ordinal: Regression Models for Ordinal Data* (version 2023.12-4). https://cran.r-project.org/web/packages/ordinal/index.html.

Cohen, J.D., D.M. Barch, C. Carter, and D. Servan-Schreiber. 1999. "Context-Processing Deficits in Schizophrenia: Converging Evidence from Three Theoretically Motivated Cognitive Tasks." *Journal of Abnormal Psychology* 108 (1): 120-33. doi:10.1037//0021-843x.108.1.120.

Conway, Andrew R.A., Michael J. Kane, Michael F. Bunting, D. Zach Hambrick, Oliver Wilhelm, and Randall W. Engle. 2005. "Working Memory Span Tasks: A Methodological Review and User's Guide." *Psychonomic Bulletin & Review* 12 (5): 769-86. doi:10.3758/BF03196772.

Cordonier, Natacha, Marion Fossard, and Maud Champagne-Lavau. 2020. "Differential Impairments in Irony Comprehension in Brain-Damaged Individuals: Insight from Contextual Processing, Theory of Mind, and Executive Functions." *Neuropsychology* 34 (7): 750-63. doi:10.1037/neu0000682.

Craik, Fergus I. M., and Ellen Bialystok. 2006. “Cognition through the Lifespan: Mechanisms of Change.” *Trends in Cognitive Sciences* 10 (3): 131-38. doi:10.1016/j.tics.2006.01.007.

Cummins, Chris. 2017. “Contextual Causes of Implicature Failure.” *Discourse Processes* 54 (3): 207-18. doi:10.1080/0163853X.2016.1142331.

De Neys, Wim, and Walter Schaeken. 2007. “When People Are More Logical under Cognitive Load: Dual Task Impact on Scalar Implicature.” *Experimental Psychology* 54 (2): 128-33. doi:10.1027/1618-3169.54.2.128.

Degen, Judith. 2013. “Alternatives in Pragmatic Reasoning.” Ph.D., New York: University of Rochester. https://www.proquest.com/docview/1465060224/abstract/32ECC29BBCE44581PQ/1.

Degen, Judith, and Michael K. Tanenhaus. 2015. “Processing Scalar Implicature: A Constraint-Based Approach.” *Cognitive Science* 39 (4): 667-710. doi:10.1111/cogs.12171.

Dieussaert, Kristien, Suzanne Verkerk, Ellen Gillard, and Walter Schaeken. 2011. “Some Effort for Some: Further Evidence That Scalar Implicatures Are Effortful.” *Quarterly Journal of Experimental Psychology*. <https://journals.sagepub.com/doi/10.1080/17470218.2011.588799>.

Grice, P. 1989. *Studies in the Way of Words*. Harvard University Press.

Hasher, Lynn, and Rose T. Zacks. 1988. “Working Memory, Comprehension, and Aging: A Review and a New View.” In *Psychology of Learning and Motivation*, edited by Gordon H. Bower, 22:193-225. Academic Press. doi:10.1016/S0079-7421(08)60041-9.

Henderson, Dori, Andrew B. Poppe, Deanna M. Barch, Cameron S. Carter, James M. Gold, John D. Ragland, Steven M. Silverstein, Milton E. Strauss, and Angus W. MacDonald. 2012. “Optimization of a Goal Maintenance Task for Use in Clinical Applications.” *Schizophrenia Bulletin* 38 (1): 104–13. doi:10.1093/schbul/sbr172.

Horn, Laurence R. 1972. “On the Semantic Properties of Logical Operators in English.” Ph.D., United States -- California: University of California, Los Angeles. https://www.proquest.com/docview/302676533/citation/1C2F7B7705574FCDPQ/1.

Horn, Laurence R. 2006. “Implicature.” In *The Handbook of Pragmatics*, 2–28. John Wiley & Sons, Ltd. doi:10.1002/9780470756959.ch1.

Huang, Yi Ting, and Jesse Snedeker. 2018. “*Some* Inferences Still Take Time: Prosody, Predictability, and the Speed of Scalar Implicatures.” *Cognitive Psychology* 102 (May): 105–26. doi:10.1016/j.cogpsych.2018.01.004.

Jones, Jessica A. H., Scott R. Sponheim, and Angus W. MacDonald III. 2010. “The Dot Pattern Expectancy Task: Reliability and Replication of Deficits in Schizophrenia.” *Psychological Assessment* 22 (1): 131–41. doi:10.1037/a0017828.

Kane, Michael J., and Randall W. Engle. 2003. “Working-Memory Capacity and the Control of Attention: The Contributions of Goal Neglect, Response Competition, and Task Set to Stroop Interference.” *Journal of Experimental Psychology: General* 132 (1): 47–70. doi:10.1037/0096-3445.132.1.47.

Katsos, Napoleon, and Chris Cummins. 2010. “Pragmatics: From Theory to Experiment and Back Again.” Language and Linguistics Compass 4 (5): 282–95. doi:10.1111/j.1749-818X.2010.00203.x.

Levinson, Stephen C. 2000. Presumptive Meanings: The Theory of Generalized Conversational Implicature. The MIT Press. doi:10.7551/mitpress/5526.001.0001.

MacDonald III, Angus W., Vina M. Goghari, Brian M. Hicks, Janine D. Flory, Cameron S. Carter, and Stephen B. Manuck. 2005. “A Convergent-Divergent Approach to Context Processing, General Intellectual Functioning, and the Genetic Liability to Schizophrenia.” Neuropsychology 19 (6): 814–21. doi:10.1037/0894-4105.19.6.814.

Marty, Paul Pierre, and Emmanuel Chemla. 2013. “Scalar Implicatures: Working Memory and a Comparison with Only.” Frontiers in Psychology 4 (July). doi:10.3389/fpsyg.2013.00403.

Marty, Paul, Emmanuel Chemla, and Benjamin Spector. 2013. “Interpreting Numerals and Scalar Items under Memory Load.” Lingua 133 (September): 152–63. doi:10.1016/j.lingua.2013.03.006.

Mashal, Nira, Gavrieli ,Ronit, and Gitit and Kavé. 2011. “Age-Related Changes in the Appreciation of Novel Metaphoric Semantic Relations.” Aging, Neuropsychology, and Cognition 18 (5): 527–43. doi:10.1080/13825585.2011.575202.

Mazzaggio, Greta, Alessandra Zappoli, and Diana Mazzarella. 2023. "Verbal Irony and the Implicitness of the Echo: The Processing of Young and Older Adults." *Pragmatics & Cognition* 30 (2): 412–43. doi:10.1075/pc.00036.maz.

Noveck, Ira. 2018. Experimental Pragmatics: The Making of a Cognitive Science. Cambridge: Cambridge University Press. doi:10.1017/9781316027073.

Noveck, Ira, and Dan Sperber. 2012. "The Why and How of Experimental Pragmatics: The Case of 'Scalar Inferences.'" In Meaning and Relevance, edited by Dan Sperber and Deirdre Wilson, 307–30. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139028370.018.

Phillips, Louise H., Roy Allen, Rebecca Bull, Alexandra Hering, Matthias Kliegel, and Shelley Channon. 2015. "Older Adults Have Difficulty in Decoding Sarcasm." Developmental Psychology 51 (12): 1840–52. doi:10.1037/dev0000063.

Pickering, Martin J., and Simon Garrod. 2004. "Toward a Mechanistic Psychology of Dialogue." Behavioral and Brain Sciences 27 (2): 169-90. doi:10.1017/S0140525X04000056.

Politzer-Ahles, Stephen, and Robert Fiorentino. 2013. "The Realization of Scalar Inferences: Context Sensitivity without Processing Cost." PLOS ONE 8 (5): e63943. doi:10.1371/journal.pone.0063943.

Politzer-Ahles, Stephen, and E. Matthew Husband. 2018. "Eye Movement Evidence for Context-Sensitive Derivation of Scalar Inferences." Edited by Rolf Zwaan and Roel Willems. Collabra: Psychology 4 (1): 3. doi:10.1525/collabra.100.

Psychology Software Tools, Inc. [E-Prime Go]. 2021. Retrieved from https://support.pstnet.com/.

Ratcliff, Roger. 1993. "Methods for Dealing with Reaction Time Outliers." *Psychological Bulletin* 114 (3): 510-32. doi:10.1037/0033-2909.114.3.510.

Roberts, Craige. 2006. "Context in Dynamic Interpretation." In *The Handbook of Pragmatics*, 197-220. John Wiley & Sons, Ltd. doi:10.1002/9780470756959.ch9.

Roberts, Craige. 2012. "Information Structure: Towards an Integrated Formal Theory of Pragmatics." Semantics and Pragmatics 6: 1-69. doi:10.3765/sp.5.6.

Ronai, Eszter, and Ming Xiang. 2021. "Pragmatic Inferences Are QUD-Sensitive: An Experimental Study." *Journal of Linguistics* 57 (4): 841-70. doi:10.1017/S0022226720000389.

Ronderos, Camilo R., and Ira Noveck. 2023. "Slowdowns in Scalar Implicature Processing: Isolating the Intention-Reading Costs in the Bott & Noveck Task." *Cognition* 238: 105480. doi:10.1016/j.cognition.2023.105480.

Scarpina, Federica, and Sofia Tagini. 2017. "The Stroop Color and Word Test." *Frontiers in Psychology* 8: 557. doi:10.3389/fpsyg.2017.00557.

Schmitt, Hannah, Maren C. Wolff, Nicola K. Ferdinand, and Jutta Kray. 2014. "Age Differences in the Processing of Context Information." *Journal of Psychophysiology* 28 (3): 202-14. doi:10.1027/0269-8803/a000126.

Servan-Schreiber, David, Jonathan D. Cohen, and Sandra Steingard. 1996. "Schizophrenic Deficits in the Processing of Context: A Test of a Theoretical Model." *Archives of General Psychiatry* 53 (12): 1105-12. doi:10.1001/archpsyc.1996.01830120037008.

Sperber, D., & Wilson, D. 1995. *Relevance: Communication and Cognition*. Oxford: Basil Blackwell.

Sprouse, Jon, and Diogo Almeida. 2013. "The Role of Experimental Syntax in an Integrated Cognitive Science of Language." In *The Cambridge Handbook of Biolinguistics*, edited by Cedric Boeckx and Kleanthes K. Grohmann, 181-202. Cambridge Handbooks in Language and Linguistics. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511980435.013.

Starr, Glenn, and Jacee Cho. 2022. "QUD Sensitivity in the Computation of Scalar Implicatures in Second Language Acquisition." *Language Acquisition* 29 (2): 182-97. doi:10.1080/10489223.2021.1990930.

Stroop, J.R. 1935. "Studies of Interference in Serial Verbal Reactions." *Journal of Experimental Psychology* 18 (6): 643-62. doi:10.1037/h0054651.

Swets, John A., and Susan T. Sewall. 1963. "Invariance of Signal Detectability over Stages of Practice and Levels of Motivation." *Journal of Experimental Psychology* 66 (2): 120-26. doi:10.1037/h0049098.

Umanath, Sharda, and Elizabeth J. Marsh. 2014. “Understanding How Prior Knowledge Influences Memory in Older Adults.” *Perspectives on Psychological Science* 9 (4): 408–26. doi:10.1177/1745691614535933.

Wilson, Deirdre, and Dan Sperber. 2006. “Relevance Theory.” In *The Handbook of Pragmatics*, 606-32. John Wiley & Sons, Ltd. doi:10.1002/9780470756959.ch27.

Yang, Xiao, Utako Minai, and Robert Fiorentino. 2018. “Context-Sensitivity and Individual Differences in the Derivation of Scalar Implicature.” *Frontiers in Psychology* 9. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01720.