Title
Time-pressure Does Not Alter the Bias Towards Canonical Interpretation of Quantifiers

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Potthoff, Ruben
Ramotowska, Sonia
Szymanik, Jakub
et al.

Publication Date
2023

Peer reviewed

Time-pressure Does Not Alter the Bias Towards the Canonical Interpretation of Quantifiers

Ruben D. Potthoff (r.potthoff@nki.nl)
Division of Psychosocial Research and Epidemiology, Netherlands Cancer Institute, Plesmanlaan 121
Amsterdam, 1066 CX The Netherlands &
Department of Experimental Psychology, Utrecht University, Heidelberglaan 8
Utrecht, 3584 CS The Netherlands

Sonia Ramotowska (sonia.ramotowska@hhu.de)
Department of Linguistics, Heinrich Heine University, Universitätsstraße 1
Düsseldorf, 40225 Germany

Jakub Szymanik (jakub.szymanik@gmail.com)
Center for Brain/Mind Sciences and the Computer Science Department, University of Trento, Corso Bettini 31
Rovereto TN, 38068 Italy

Leendert van Maanen (l.vanmaanen@uu.nl)
Department of Experimental Psychology, Utrecht University, Heidelberglaan 8
Utrecht, 3584 CS The Netherlands

Abstract

The Interface Transparency Thesis (ITT) proposes that people tend to use a canonical interpretation of linguistic expressions, even when this interpretation is sub-optimal for the task at hand. The current paper sought to investigate this claim further by adding a time-pressure manipulation to a quantified sentence verification task and analyzing the results through a computational model of decision-making. The results indicate that time pressure -while effectively changing behavioral responses- does not alter cognitive processes associated with quantifier verification, thus supporting the ITT.

Keywords: quantifiers; verification; speed-accuracy trade-off; diffusion decision model

Introduction

The meaning of natural language quantifiers (e.g., *more than half*, *most*, *less than half*, and *least*) can be captured in the form of truth-conditional representations, see, e.g., (Szymanik, 2016). The Interface Transparency Thesis (ITT) proposed by Lidz, Pietroski, Halberda, and Hunter (2011) claims: "Speakers exhibit a bias towards the verification procedures provided by canonical specifications of truth conditions. (p.229)" Some studies provided evidence in favor of the ITT (Lidz et al., 2011; Knowlton et al., 2021). Specifically, Lidz et al. (2011) have shown that the bias towards canonical procedures may be stronger than the bias towards the simplest strategies. The natural question appears: how strong is this bias? Some studies show that the mental representation of quantifiers might differ between different contexts (Register, Mollica, & Piantadosi, 2020), others point to individual differences (Talmina, Kochari, & Szymanik, 2017). We ask how robust the bias is under extra time pressure.

The current study investigates four Dutch quantifiers *more than half* (*meer dan de helft*), *most* (*de meeste*), *less than half* (*minder dan de helft*), and *least* (*de minste*) and aims to test the prediction of the ITT experimentally. To operationalize mental representations of quantifiers, we apply a computational model proposed by Schlotterbeck, Ramotowska, van Maanen, Szymanik, et al. (2020) and Ramotowska, Steinert-Threlkeld, van Maanen, and Szymanik (2023). To test predictions of the ITT, we tested if the linguistic biases toward specific mental representations of quantifiers would resist the time pressure manipulation.

Computational Modeling of Quantifier Verification

Ramotowska et al. (2023) investigated two English quantifiers *most* and *more than half* in a truth-value judgement task. In this task, participants read a sentence of the form "67% of the As are B", followed by a sentence of the form "Most / more than half of the As are B". They verified the second sentence based on the information from the first sentence. To jointly analyze participants' responses and reaction times, Ramotowska et al. (2023) fitted a modification of the Diffusion Decision Model (DDM, Ratcliff & McKoon, 2008). They operationalized mental representations of quantifiers in terms of two model parameters: threshold (proportion above which participants judge sentences as true) and vagueness (increased uncertainty around the threshold indicated by slower responses).

The DDM, applied by Schlotterbeck et al. (2020) and Ramotowska et al. (2023), is a cognitive model for two-alternative forced choices that describes the decision formation process as the accumulation of evidence towards one of two decision boundaries (Ratcliff & McKoon, 2008). As can be seen in Figure 1A, two boundaries represent the response alternatives, and the distance between them (denoted by a) represents the amount of evidence required before a decision-maker commits to a response. Once the accumulation process (represented by v or the blue arrows in the figure) crosses a boundary, decision formation is ended, and a choice is made.

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

The time required to reach this point constitutes the decision process. While there are many parameters in the DDM, the key components for the current study are the decision boundary parameter a and drift rate parameter v.

The drift rate is the key parameter of the DDM that captures the evidence accumulation process. Because the speed of evidence accumulation in the sentence-verification task depends on the proportion, Schlotterbeck et al. (2020) and Ramotowska et al. (2023) operationalized the drift rate parameters as a logistic function of that proportion (Figure 1B):

$$v(p) = v_L + \frac{v_U - v_L}{1 + e^{B(p-M)}},$$

Where p is the specific percentage for which the drift rate is being predicted, M is the midpoint of the logistic curve, corresponding to the threshold, B is the growth rate of the logistic curve, corresponding to vagueness, v_L are v_U the lower and upper drift rate asymptotes, respectively.

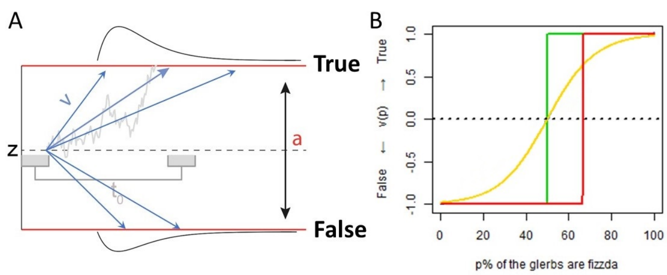

Figure 1: A. A Diffusion Decision Model representation of choice between a *True* and a *False* sentence verification. On each trial (verification decision), the decision formation process is conceptualized as a random walk (grey line) with average direction v (drift rate, represented by the blue arrows) until one of two boundaries is crossed (separated by the amount of evidence a). The accumulation process starts at the starting point z, and the total response time RT for the model is the sum of the decision time and the encoding and motor execution time, together referred to as t_0. B. The drift rate $v(p)$ is modeled as a logistic function of the proportion (p%) that participants are asked to verify. The proportion (p%) for which the drift rate ($v(p)$) is zero is the quantifier's threshold. The shape of the function indicates the quantifier's vagueness. Shown here are three example drift rate-proportion relations: the green line is a drift rate for non-vague quantifiers with a threshold at 50%, the red line is a drift rate for non-vague quantifiers with a threshold above 50%, the yellow line is a drift rate for vague quantifiers with the threshold at 50%.

The flexibility of this drift rate function allows to account for different effects on verification times and choices. Firstly, the threshold of a quantifier (represented here by M) is conceptualized as the proportion for which the drift rate is zero. For quantifiers such as *more than half* and *less than half*, this should logically be at 50% (see the green line in Figure 1B). A higher threshold above which individuals respond that a quantified sentence is true given a particular percentage can be expressed as a shift in the midpoint (compare red and green lines in Figure 1B). Secondly, more uncertainty around the threshold can be modeled with different growth rates (compare the yellow line in Figure 1B relative to the green line). Using this approach, Ramotowska et al. (2023) found variability in midpoints between participants for *most* but not for *more than half*, suggesting different thresholds for these quantifiers, and a difference in mean growth rate, which was taken as an indication for a different verification process.

To sum up, the specification of the drift rate as a logistic function makes it possible to measure two crucial aspects of quantifier representation - the threshold and vagueness. If these parameters reflect the *canonical specification* of the quantifier's truth conditions, they should not change when participants verify quantifiers under time pressure. This is because the time pressure, while affecting participants' behavior, does not affect the specification of truth conditions.

Speed-accuracy Trade-off Manipulation

The current study aims to understand whether the cognitive processes associated with quantifier verification are the same under time pressure or whether they categorically differ. To answer this question, we introduced a speed-accuracy trade-off manipulation. The speed-accuracy trade-off has been well-known within psychology for over a century (Woodworth, 1899). It refers to the fact that one can increase the speed of their reaction at the cost of accuracy and increase accuracy at the cost of speed.

In one linguistic study, the speed-accuracy trade-off was successfully applied to investigate the verification of a quantifier *some* that has two competing interpretations: lower-bound (*some and possibly all*) and upper-bound (*only some*, Bott & Noveck, 2004). Bott and Noveck (2004) found that under time pressure, participants more often interpreted *some* with its lower-bound meaning than with its enriched, upper-bound meaning. They concluded that under time pressure, participants do not compute enriched meaning because this is cognitively costly. In non-linguistic domains, the speed-accuracy trade-off has also been a successful method to understand whether cognitive strategies can be flexibly adjusted in the face of time pressure. For example, when making value-based judgments between two lotteries, time pressure leads to a change of strategy, where individuals tend to follow their natural default behavior rather than making an optimal choice (Couto, Van Maanen, & Lebreton, 2020). In contrast, in perceptual decision-making tasks, the dominant observation is that individuals do not change their cognitive strategy for making a decision, despite differences in observed behavior. That is, when placed under time pressure, participants respond faster but also make more erroneous responses. However, the consensus is that this behavioral pattern is the consequence of executing the same cognitive strategy, but for a shorter period of time (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010; Katsimpokis, Hawkins, & van Maanen, 2020; Heitz, 2014; Van Maanen, 2016).

An often-used model to investigate the effects of the speed-accuracy trade-off on various other underlying cognitive processes is the aforementioned DDM (Ratcliff & McKoon, 2008). The DDM captures the effect of time pressure through its boundary separation parameter a. For example, Winkel et al. (2012) asked participants to make two-alternative perceptual choices and asked them to either focus on being fast or on being accurate. They found that the best explanation for the observed reaction time and accuracy differences was a difference in boundary separation a. The intuition behind this and similar findings (van Maanen et al., 2011; Katsimpokis et al., 2020; Forstmann et al., 2008; Boehm, Van Maanen, Forstmann, & Van Rijn, 2014) is that a shorter distance to a boundary provides a faster response, as desired under speed stress. At the same time, due to the noisiness of the evidence accumulation, a lower boundary separation also means an increased probability of incorrect response, yielding the speed-accuracy trade-off.

The main contribution of this paper is the introduction of time pressure manipulation to test its effect on the canonical representation of quantifiers. In the spirit of the ITT, the linguistic biases associated with specific quantifiers should resist time pressure. Hence, the verification of quantifiers measured by parameters of the drift rate should stay the same in speed and accuracy conditions. The only parameter that may change is the boundary separation parameter, which is not related to quantifier meaning representation (Schlotterbeck et al., 2020). This hypothesis entails that behavioral differences that we observed due to time pressure can be solely attributed to the time pressure manipulation, and the cognitive processes underlying verification remain the same. This means that similar to what is often observed in perceptual decision-making, but contrary to the Bott and Noveck (2004) study, time pressure only changes the boundary separation parameter and not any other parameter of the DDM.

Methods

Materials & Participants

The experiment was conducted online and was approved by the local ethics committee. The experiment included a digital consent form hosted through the Qualtrics survey tool, which referred the participant to the actual experiment (*Qualtrics*, 2021). The experiment used the experiment hosting service MindProbe to host the experiment itself (*MindProbe*, 2021). MindProbe hosts experiments that integrate Just Another Tool for Online Studies (JATOS), which our experiment integrates (*JATOS*, 2021). The experiment itself is programmed using JavaScript, and all analyses were done in R.

We recruited 48 participants (38 female) for this experiment, 5 of which were excluded because they indicated that they had not followed instructions. To exclude fast-guessing participants, we used as an exclusion criterion that participants, on average, needed to be slower than 300ms (Ramotowska, Steinert-Threlkeld, Van Maanen, & Szymanik, 2020; Ramotowska et al., 2023; Schlotterbeck et al., 2020). However, in our experiment, no participants had to be excluded for this reason. Participants violating the expected monotonicity effect (more 'true' responses for *most/more than half* if the proportion increases above 50%, and the opposite for *least/less than half*) would also be excluded (Schlotterbeck et al., 2020), yet no such participants were present either. Finally, all trials with response times larger than 5 seconds or shorter than 100ms have also been excluded (1.08% of the trials).

Design

The trial setup followed a similar setup as found in Ramotowska et al. (2023). In a trial, participants were given a sentence of the formula "X of the A's are B" ("X van de A's zijn B" in Dutch), where X is one of the quantifiers *most* (de meeste), *more than half* (meer dan de helft), *least* (de minste) and *fewer than half* (minder dan de helft). These quantifiers were chosen based on their prevalence in related research. A and B are two-syllable pseudowords, with 80 pseudo-adjectives and 80 pseudo-nouns, adapted from randomly generated words generated through pseudo-word generation software Wuggy (Keuleers & Brysbaert, 2010). The pseudowords were randomly paired for each trail. Following the first sentence, the participants were presented with another sentence of the form "x% of the A's are B's" (x% van de A's zijn B). Here, x% is a random percentage between 1% and 99%. There were no trials with 0% or 100%, as there are no clear upper limits for *most* (Ariel, 2003). The participants had to judge whether the initial sentence accurately described the second sentence by pressing one of two specified buttons on their keyboard (J for false, L for true).

The experiment had two conditions, one in which the participants had to focus on responding quickly while they had to focus on responding accurately in the other. Each condition consisted of 75 trials, and each was performed twice for 300 trials. As each participant had to perform both conditions, the experiment followed a within-participant design. The order of the conditions was counterbalanced between subjects, but due to the number of excluded participants, there was a slight imbalance (19 participants did 'Fast>Slow>Fast>Slow' while 24 did 'Slow>Fast>Slow>Fast'). Percentages and quantifiers were randomized for each trial.

Before each condition, instruction was given to either focus on responding fast or on focusing on responding accurately. As motivation to respond quickly, participants were told that an image of a cute puppy would be displayed at the end of the experiment if they properly followed the instruction. In the end, participants were shown a picture of a cute puppy regardless of whether they had followed the instructions.

Diffusion Decision Model Analysis

Our specification of DDM uses 10 different parameters. In addition to the standard parameters a, t_0, z, sv, sz, st_0, we specified the drift rate by the logistic relationship with the percentage outlined in the introduction. This leads to an additional four parameters (v_L, v_U, B, and M). To calculate the optimal values for each parameter, we used maximum likelihood estimation (Myung, 2003). Specifically, to estimate the parameters, we used particle swarm optimization to optimize a set of parameters using parallel searches of the parameter space (Clerc, 2010). The parameters for variability *(sz, st0, eta)* were set to 0 for each of the models, as estimates for these parameters are inherently unreliable and poor estimates for variability could negatively affect the estimation of the other parameters (Boehm et al., 2018). To be able to fit the same model on the negative quantifiers (*least, less than half*) as on the positive quantifiers (*most, more than half*), we flipped the true and false responses for the negative quantifiers in data preprocessing.

Our initial model specification allowed each parameter to differ between speed-stress conditions and quantifiers. We simplified this most complex model via stepwise deletion of factors by constraining the parameter that was least likely to differ between conditions or quantifiers to the same value for those conditions/quantifiers, according to an analysis of variance (ANOVA). This procedure was stopped when ANOVAs for each parameter suggested that the parameter significantly differed between the remaining factors. As a negative control of the experimental manipulation, we also constrained a over speed-stress conditions. The constrained models were compared using Akaike Information Criterion (AIC) (Akaike, 1974). AIC penalizes more complex models, as these necessarily have better goodness of fit than more constrained models (Pitt & Myung, 2002).

To draw inferences from the DDM parameters, we computed Bayesian Model Averages (BMA) for all parameters (Hinne, Gronau, van den Bergh, & Wagenmakers, 2020). Specifically, a weighted average of the parameters was computed based on all constrained models, where the models were weighted according to their Akaike weight (Wagenmakers & Farrell, 2004).

Results

Descriptive Statistics

The mean response times in the experiment are displayed in Figure 2. As can be seen, the response times for the speed stress condition were lower than the response times for the accuracy condition for all quantifiers. We substantiate this observation using linear mixed-effects models of the RT data, with fixed effects Quantifier, Instruction (speed or accuracy), and the percentage deviation from 50% as well as their interactions. As random effects, we included random slopes for quantifiers, instruction, and the percentage deviation from 50% for each participant. Via backward stepwise deletion of factors, we identified the regression model that best-balanced goodness-of-fit and degrees of freedom according to a likelihood ratio test. This model included the main effects of Quantifier, Instruction, and Percentage as well as the Quantifier $\times$ Percentage interaction. As a manipulation check, we found a main effect of speed-accuracy trade-off instruction, in that accuracy-stressed trials were indeed verified slower ($\beta_{SAT} = 0.138; p < 0.001$). The analysis also revealed that compared to *less than half*, *most* and *more than half* are verified faster ($\beta_{Most} = -0.185; p < 0.001$ and $\beta_{MTH} = -0.202; p < 0.001$), which is consistent with the monotonicity effect (Schlotterbeck et al., 2020). However, there was no main effect of *least*. Crucial for our main research question, the best model after stepwise deletion of factors did not include any interaction with speed-accuracy trade-off instruction. The quantifiers *most*, *more than half*, and *least* showed interactions with the percentage mentioned in the first sentence ($\beta_{Most\times p} = -0.092; p < 0.001$, $\beta_{MTH\times p} = -0.083; p < 0.001$, $\beta_{Least\times p} = -0.041; p = 0.0027$). For these quantifiers, we found that when percentages were closer to 50%, RTs were higher. The same interaction was not observed for *less than half*.

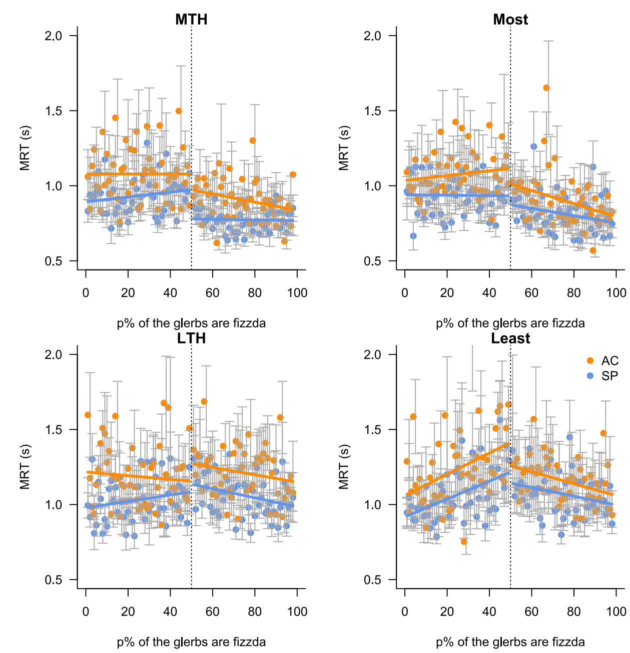

Figure 2: Mean response times (MRT) in seconds of all quantifiers show a clear speed-accuracy trade-off effect. Error bars indicate standard errors of the mean. Regression lines illustrate trends. SP: speed instruction; AC: accuracy instruction, MTH: *more than half*, LTH: *less than half*

Diffusion Decision Model Analysis Results

For the following comparisons of model fit, the best fitting model according to AIC was used, as discussed in the 'Diffusion Decision Model Analysis' section. This model constrained all parameters except for decision boundary a. This model fits the observed data well, as seen in Figure 3. The exception may be the quantifier *least*, where the model slightly underestimates the proportion of *true* responses.

Model Constraints Beyond setting the variability parameters, *sz, st0 and eta*, to 0 for each model due to their inherent unreliability, further stepwise deletion of factors led to

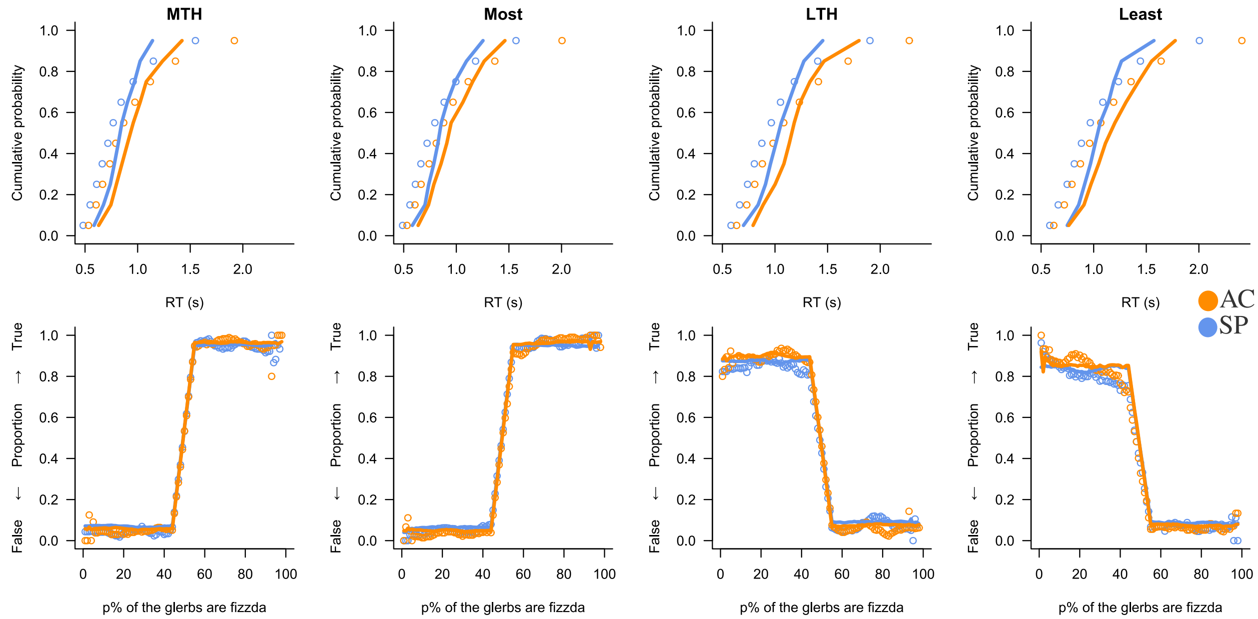

Figure 3: The overall most preferred model captures meaningful variability in the behavioral data. Top row: Averaged (Vincentized, Ratcliff, 1979), cumulative response time distribution of the data (points), with the model predictions overlaid. Bottom row: Running average of the proportion of *true* and *false* responses in the data (points), with the same running average of the model predictions overlaid. SP: speed instruction; AC: accuracy instruction; MTH: *more than half*; LTH: *less than half*

the following constraints: The midpoint was fixed at $M = 0$. $M = 0$ effectively ties the threshold to the logical interpretation of the quantifiers, where a change in truth value is also linked to 50%. The growth rate B was set to $B = \inf$, which effectively models a step function and eliminates vagueness (Ramotowska et al., 2023). The non-decision time t_0, starting point z, and the upper and lower drift rate asymptotes were all set equal across speed/accuracy conditions, indicating that participants do not change these parameters time under speed stress.

Early analysis of constrained models indicated that constraints over most parameters improved the AIC. The best-fitting models were those in which nearly all parameters are either constrained across conditions or constrained to an exact value. The only parameter over which constraints only worsened AIC was the boundary separation parameter a. This suggests that the only change in the decision-making process under time stress is a reduction in the decision boundary.

To consider, however, that not every model is equally preferred by each individual, we computed weighted averages of each model parameter according to AIC. Figure 4 presents the distribution of these weighted parameter values per quantifier and speed-stress condition. ANOVA tests performed on these model averages show that the only parameter on which time pressure had a significant effect was the boundary separation parameter a, $F(1,42)=17.32$, $p < .001$.

Discussion

To summarise, we found clear evidence for the speed-accuracy trade-off in the truth value judgment task. When put under time pressure, participants responded faster than in the accuracy condition. In terms of model parameters, the only parameter that was affected by time pressure was the boundary separation parameter a. Crucially, however, the time pressure manipulation did not affect the linguistic behavior of participants. In particular, the meaning representations of quantifiers captured by the drift rate parameters of the model resisted the time pressure manipulation.

The fact that cognitive processes associated with meaning representation do not differ under time pressure supports the ITT, as we hypothesized. According to the ITT, each quantifier has a preferred mental representation that guides the verification process. In addition, to support the ITT, our study shows the usefulness of speed-accuracy manipulation and the DDM application in the investigation of linguistic processing. For example, previous studies (Bott & Noveck, 2004) showed that scalar implicatures are less frequently derived under time pressure. This has been interpreted as evidence for a delay of pragmatic processing compared to semantic processing of meaning, yet these studies are burdened by a confound between the change in participants' behavior due to changes in linguistic representation and due to time pressure manipulation. By mapping the semantic and pragmatic processing on different model parameters than the effect of time pressure, we could deal with this confound.

Acknowledgments

Both the experimental data and the analyses performed in this paper can be found on the Open Science Framework page for this paper at `https://osf.io/6xkbs/`.

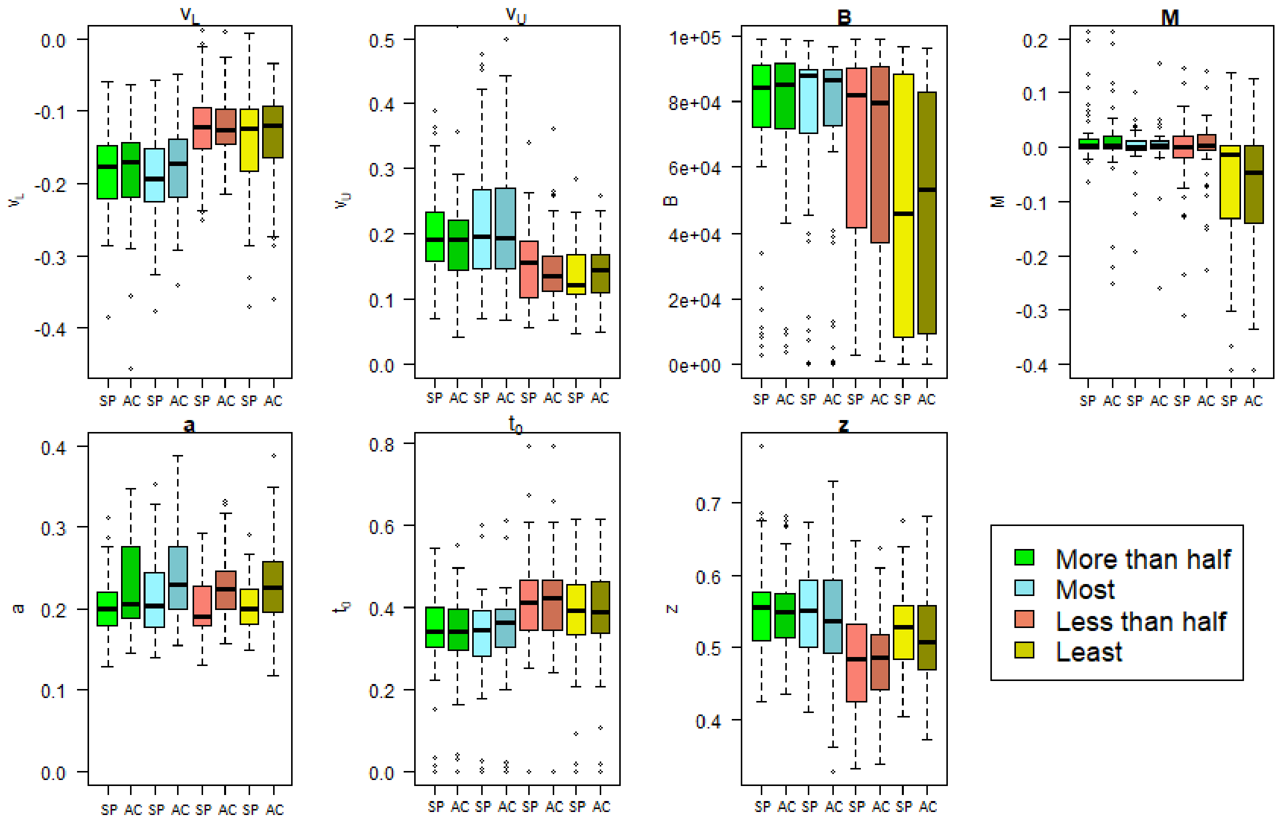

Figure 4: Box plots displaying the parameter values for each parameter across all conditions. On the x-axis, the speed stress condition is displayed: SP: speed instruction; AC: accuracy instruction. The quantifiers are indicated with different colors, from left to right: *least* in green, *less than half* in blue, *most* in pink, and *more than half* in yellow. The speed stress condition is also consistently displayed in a brighter shade of color for visual clarity. The y-axis represents the value for the parameter.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716–723.

Ariel, M. (2003). Does most mean'more than half'? In *Annual meeting of the berkeley linguistics society* (Vol. 29, pp. 17–30).

Boehm, U., Annis, J., Frank, M. J., Hawkins, G. E., Heathcote, A., Kellen, D., ... others (2018). Estimating across-trial variability parameters of the diffusion decision model: Expert advice and recommendations. *Journal of Mathematical Psychology*, *87*, 46–75.

Boehm, U., Van Maanen, L., Forstmann, B. U., & Van Rijn, H. (2014). Trial-by-trial fluctuations in CNV amplitude reflect anticipatory adjustment of response caution. *NeuroImage*, *96*, 95–105. doi: 10.1016/j.neuroimage.2014.03.063

Bogacz, R., Wagenmakers, E.-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed–accuracy tradeoff. *Trends in neurosciences*, *33*(1), 10–16.

Bott, L., & Noveck, I. A. (2004). Some utterances are underinformative: The onset and time course of scalar inferences. *Journal of memory and language*, *51*(3), 437–457.

Clerc, M. (2010). *Particle swarm optimization* (Vol. 93). John Wiley & Sons.

Couto, J., Van Maanen, L., & Lebreton, M. (2020). Investigating the origin and consequences of endogenous default options in repeated economic choices. *PloS one*, *15*(8), e0232385.

Forstmann, B. U., Dutilh, G., Brown, S. D., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 17538–17542. (ISBN: 0027-8424 Place: Forstmann, BU Univ Amsterdam, Dept Psychol, Amsterdam Ctr Study Adapt Control Brain & Behav, NL-1018 WB Amsterdam, Netherlands Univ Amsterdam, Dept Psychol, Amsterdam Ctr Study Adapt Control Brain & Behav, NL-1018 WB Amsterdam, Netherlands Univ Amsterdam,)

Heitz, R. P. (2014). The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Frontiers in neuroscience*, *8*, 150.

Hinne, M., Gronau, Q. F., van den Bergh, D., & Wagenmakers, E.-J. (2020). A conceptual introduction to bayesian model averaging. *Advances in Methods and Practices in Psychological Science*, *3*(2), 200–215.

Jatos. (2021). `https://www.jatos.org/`.

Katsimpokis, D., Hawkins, G. E., & van Maanen, L. (2020). Not all speed-accuracy trade-off manipulations have the same psychological effect. *Computational Brain & Behavior*, *3*(3), 252–268.

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior research methods*, *42*(3), 627–633.

Knowlton, T., Hunter, T., Odic, D., Wellwood, A., Halberda, J., Pietroski, P., & Lidz, J. (2021). Linguistic meanings as cognitive instructions. *Annals of the New York Academy of Sciences*, *1500*(1), 134–144.

Lidz, J., Pietroski, P., Halberda, J., & Hunter, T. (2011). Interface transparency and the psychosemantics of most. *Natural Language Semantics*, *19*(3), 227–256.

Mindprobe. (2021). `https://mindprobe.eu/`.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of mathematical Psychology*, *47*(1), 90–100.

Pitt, M. A., & Myung, I. J. (2002). When a good fit can be bad. *Trends in cognitive sciences*, *6*(10), 421–425.

Qualtrics. (2021). `https://www.qualtrics.com/`.

Ramotowska, S., Steinert-Threlkeld, S., Van Maanen, L., & Szymanik, J. (2020). Most, but not more than half, is proportion-dependent and sensitive to individual differences. In *Proceedings of sinn und bedeutung* (Vol. 24, pp. 165–182).

Ramotowska, S., Steinert-Threlkeld, S., van Maanen, L., & Szymanik, J. (2023). Uncovering the structure of semantic representations using a computational model of decision-making. *Cognitive Science*, *47*(1), e13234.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological bulletin*, *86*(3), 446.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural computation*, *20*(4), 873–922.

Register, J., Mollica, F., & Piantadosi, S. T. (2020). *Semantic verification is flexible and sensitive to context.*

Schlotterbeck, F., Ramotowska, S., van Maanen, L., Szymanik, J., et al. (2020). Representational complexity and pragmatics cause the monotonicity effect. In *Cogsci.*

Szymanik, J. (2016). *Quantifiers and cognition. logical and computational perspectives*. Springer.

Talmina, N., Kochari, A., & Szymanik, J. (2017). Quantifiers and verification strategies: connecting the dots. In *Proceedings of the 21st Amsterdam Colloquium.*

Van Maanen, L. (2016). Is there evidence for a mixture of processes in speed-accuracy trade-off behavior? *Topics in Cognitive Science*, *8*(1). doi: 10.1111/tops.12182

van Maanen, L., Brown, S. D., Eichele, T., Wagenmakers, E.-J., Ho, T., Serences, J., & Forstmann, B. U. (2011). Neural correlates of trial-to-trial fluctuations in response caution. *Journal of Neuroscience*, *31*(48), 17488–17495.

Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*, 192–196.

Winkel, J., Van Maanen, L., Ratcliff, R., der Schaaf, M. E., Van Schouwenburg, M. R., Cools, R., & Forstmann, B. U. (2012). Bromocriptine does not alter speed-accuracy trade-off. *Frontiers in Decision Neuroscience*, *6*, Article 126.

Woodworth, R. S. (1899). Accuracy of voluntary movement. *The Psychological Review: Monograph Supplements*, *3*(3), i.