Most, but not more than half, is proportion-dependent and sensitive to individual differences

Sonia Ramotowska[1], Shane Steinert-Threlkeld[2], Leendert Van Maanen[3], Jakub Szymanik[1]

[1] Institute for Logic, Language and Computation, University of Amsterdam [2] Department of Linguistics, University of Washington
[3] Department of Psychology, University of Amsterdam

Introduction The verification strategies of quantifiers have become a central topic in semantics. According to H09's analysis, *most* (M) and *more than half* (MTH) are truth-conditionally equivalent expressions but have different logical forms and verification strategies. The logical form of M is (1) $Q(A)(B) = 1$ iff $|A \cap B| > |A - B|$, while that of MTH is (2) $Q(A)(B) = 1$ iff $|A \cap B| > 1/2|A|$. MTH triggers a verification strategy that refers to the threshold, namely half. A verification strategy for M, in turn, requires computation of the target set (As that are B) and the set complement (As that are not B).

Additionally, S16 linked the logical form of MTH (2) with the precise number system and the logical form of M (1) with the Approximate Number System (ANS) (S16, D97). As a consequence, the meaning of M implies a "significantly more than half" interpretation. Moreover, P09 and L11 provided empirical evidence that the standard ANS model fits data on the verification of M very well. The association between ANS and the verification strategy of M results in proportion-dependent performance: as the proportion of As that are B (compared to those that are not B) grows, accuracy increases.

To summarize, the existing findings suggest that: M and MTH have different verification strategies (H09, S16), the verification of M is ratio-dependent (P09) and the representation of M ranges between the truth-conditionally correct, literal "MTH" reading, and a preferred "significantly MTH" interpretation (S16). This ambiguity of representation of M suggests that it will be more sensitive to individual differences (IDs), e.g. the choice of threshold, than MTH. However, the effect of IDs in the representation of M and MTH on the verification process is largely unexplored (K18).

We aimed to address this gap and test differences in verification strategies of M and MTH taking into account IDs in the representations of these quantifiers - participants' individual thresholds (ITs). In the current experiment, participants had to evaluate a sentence with a quantifier (M or MTH) based on a sentence providing the proportion of As that are B. We analyzed participants' ITs and their effect on the speed of verification process. We fitted an evidence accumulation model: the Diffusion Decision Model (DDM; R08). The DDM represents two-choice decisions (e.g. true/false) as an evidence accumulation process (with a drift-rate parameter v), which starts in one point (z parameter) and finishes when enough evidence is accumulated in one of the directions (boundary parameters a). Additionally, DDM has a non-decision time parameter Ter (the time required to execute the response after the decision is made). Because the DDM accounts for both responses and reaction times (RTs) distributions, it can be used to model individual choices and their effect on the speed of decision making. Moreover, recently, the DDM was successfully applied to ANS numerosity judgment tasks (R18).

Following H09's analysis, we hypothesized that M will be verified using strategy (1), while MTH (2). Moreover, based on P09 and S16 we assumed that M is associated with an ANS based strategy, while MTH with a precise strategy. Hence, we predicted that (H1) the verification of M, as opposed to MTH, will be dependent on proportion. Moreover, we predicted that (H2) participants will vary more in their interpretation of M than MTH and therefore that there will be a difference in the distribution of ITs of M and MTH. Finally, H09 showed that the speed of quantifier verification depends on the choice of cognitive representation of the quantifier. If the representation of M varies between participants, then also the speed of verification should vary. Therefore, we predicted that (H3) the difference in representation ITs will affect RTs only for M.

We aimed to strengthen these predictions by applying the DDM and showing that M and MTH differ in the evidence accumulation process - the drift rate parameter (H4). Additionally, we predicted a correlation between ITs and DDM threshold parameters (H5).

Methods We performed a truth-value judgment experiment and a direct replication. We used a purely linguistic paradigm to avoid non-linguistic effects. Participants had to decide if the first sentence of the form: "Q of the As are B" (with Q a quantifier: *M, MTH, fewer than half, many, few*) is true based

on the information from the second sentence of the form: "X% of the As are B" (with X% ranging between 1%-99% excluding 50%). To minimize other linguistic effects beyond quantifier meaning, we used pseudowords (K10) from English nouns (As) and adjectives (Bs). We collected participants' responses and RTs.

Results To test the H1 we ran mixed effects regression model with RTs as dependent variable and proportion (z-scored), quantifier, response (true/false), and their interactions as predictors. We included by-subject intercept and random slope for response in original experiment and additionally by-subject random slope for proportion in replication. We set true responses and MTH as baseline. We found not significant main effect of proportion ($\beta = -21.71$, $t = -.99$, $p < .32$) ($\beta = -29.01$, $t = -1.05$, $p < .29$), but significant proportion quantifier interaction ($\beta = -147.04$, $t = -4.63$, $p < .0001$) ($\beta = -154.83$, $t = -4.32$, $p < .0001$) in both original and replication experiments respectively.

In the next step, we estimated ITs for each quantifier by fitting the logistic function (f(x) = 1/(1+exp((x-IT)/k))), where x indicates the proportion in the second sentence of each trial, k the steepness of the curve and f(x) represents the probability of the true response in each trial. For those participants for whom the parameters of the model could not be estimated, we computed the ITs as the mean of the highest proportion where the response was false and the lowest proportion where the response was true. We excluded participants with ITs above 100% and below 0%. The final sample consisted of 72 participants in the original experiment and 66 in the replication (Fig. 1). The mean threshold for M was 52.62%; 51.21% and for MTH 49.49%; 49.42% in the original experiment and replication respectively. We found that the ITs distributions of M and MTH are significantly different ($D = 0.29$, $p = .004$) in original experiment and ($D = 0.27$, $p = .01$) in replication.

To understand how ITs influence RTs, we ran a mixed effect regression model with RTs as dependent variable and proportion (z-scored), ITs (z-scored), response (true/false), and their interactions as predictors. We set true responses as baseline and included by-subject random intercept for M and MTH and by-subject random slope for response in the original experiment and for response and proportion in the replication for M (only by-subject random slopes that improved the model were included). In the original experiment we found for M a significant main effect of ITs ($\beta = 167.74$, $t = 3.97$, $p < .0001$), proportion ($\beta = -214.85$, $t = -7.58$, $p < .0001$) and ITs x proportion interaction ($\beta = -80.69$, $t = -2.83$, $p < .001$), but no significant effects for MTH (respectively: $\beta = 20.43$, $t = 1.01$, $p = .31$; $\beta = -21.63$, $t = -1.16$, $p = .25$); $\beta = -1.65$, $t = -.12$, $p = .9$). In the replication we found for M significant main effects of proportion ($\beta = -198.34$, $t = -6.1$, $p < .0001$) and approaching significant level effect of ITs ($\beta = 82.15$; $t = 1.79$, $p = 0.08$) and no effect for MTH ($\beta = -28.20$, $t = -1.19$, $p = .23$), ($\beta = -7.99$, $t = -.23$, $p = 0.82$) (Fig 1).

Finally, we fitted the DDM to RTs and response data from the original experiment for each participant separately. We set the drift rate parameter to be dependent on proportion via the generalized logistic function: (La + (Ua - La) / (1 + exp(-B*(x-Mid)))), where *La* is the lower asymptote of the generalized logistic function (drift rate parameter *v*), *Ua* is upper asymptote, *B* is the growth rate, *x* is the proportion in the second sentence of each trial and *Mid* is the midpoint (DDM threshold parameter). We estimated the optimal parameters of the simple DDM (i.e. without variability parameters) using particle swarm optimization of the maximum likelihood of parameters given the data. We systematically constrained parameters across conditions (by assessing BIC values (S78)) to understand which cognitive modeling process explains the observed behavioral patterns best. The final model had constrained *z*, *a*, *Ter* and *La* and *Ua* parameters (Fig 1). We found significant differences between M and MTH for both the *Mid* parameter ($t = -3.76$, $p < .001$) and *B* parameter ($t = 3.52$ $p < .001$). Moreover, we found a significant correlation between IT estimated using logistic regression and threshold parameter (*Mid*) from DDM for M ($r = .88$; $p < .0001$) and MTH ($r = .35$; $p < .01$). This two correlations were not significantly different ($\beta = -.03$ $t = -.02$; $p = .98$).

Discussion In line with our hypotheses, we found the effect of proportion (H1) only for M, but not MTH. Moreover, we found differences in the distributions of ITs of M and MTH (H2). In the original experiment, we found the effects of ITs and ITs proportion interaction for M, but not MTH (H3) and in the replication the effect of ITs was close to significance level effect for M. Additionally, using the DDM model we found differences in the evidence accumulation process between M and MTH (H4)

and correlations between ITs and DDM threshold parameter (H5). These findings strongly suggest that M and MTH differ in terms of verification strategy and in terms of individual representation.

Using a novel, purely linguistic paradigm, we strengthen H09's findings that MTH is verified using a precise strategy where the given proportion is compared to half, while M is verified using a strategy where two proportions (target and complement sets) are compared. Moreover, we extended P09's finding that M is proportion-dependent by showing the effect of proportion and ITs on RTs. Because P09 did not use MTH in their experiment, they did not show that the proportion effect is exclusive for M or rule out the possibility that the effect of proportion is an artifact of the experimental paradigm. We presented results that filled this gap and strongly suggested that only M is proportion-dependent.

H09 and P09 did not consider that M and MTH can also differ in terms of ITs. We extended their findings by empirically testing the difference in threshold for M and MTH and by showing that the differences in cognitive representations of M and MTH can affect the verification process.

We used the DDM to investigate the source of differences between M and MTH. We confirmed that M and MTH differ in both ITs and quality of the representation. The modeling results provided evidence that mean threshold for M was higher than for MTH, however did not support S16 findings that M is not used for proportion close to 50%. The DDM results also pointed out the differences in growth rate parameter B. The fact that participants were less certain during the verification of M, had lower growth rate, supports the claim that M is represented in a less precise manner than MTH, in ANS.

Taken together, we confirmed that M and MTH are associated with different verification strategies by using a novel experimental paradigm and applying the evidence accumulation modeling paradigm. Our methods also allow us to trace one source of differences between M and MTH to individual differences in logical form, which then manifest themselves in verification behavior.

Reference D97: Dehaene, S. *The number sense: How the mind creates mathematics* (1997). • H09: Hackl, M. *Nat Lang Sem 17*, 63-98 (2009). • K10: Keuleers, E, Brysbaert, M. *Behav Res Methods 42(3)*, 627-633 (2010). • K18: Kidd, E. et al *Trends in Cog Sci.* 22 (2), 154-169 (2018). • L11: Lidz, J. *Nat Lang Sem 19(3)*, 227-256 (2011). • P09: Pietroski, P. et al *Mind & Language* 24 (5), 554-585 (2009). • R08: Ratcliff, R, McKoon, G. *Neural Comput 20(4)*, 873-922 (2008). • R18: Ratcliff, R, McKoon, G. *Psychol Rev. 125(2)*, 183-217 (2018). • S16: Solt, S. *Language 92*,. 65-100 (2016). • S78: Schwarz, G. *Ann. Statist. 6*, 461-464 (1978).
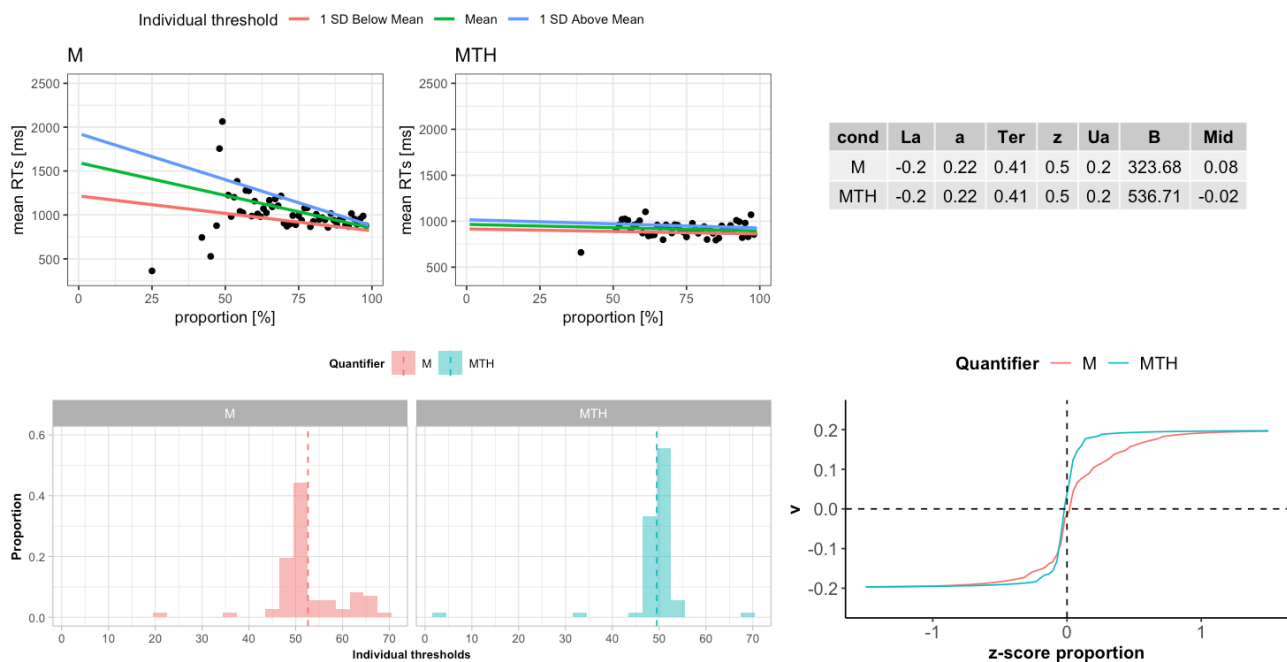

Fig. 1. (**A**) Mean RTs for M and MTH with regression lines for *mean* threshold, threshold 1*SD* above *mean* and 1*SD* below *mean* (original experiment). Only responses *true* above threshold were plotted. (**B**) Distribution of individual thresholds in original experiment for each quantifier. The dashed lines indicate mean threshold ($N = 72$): MTH = 49.49%; M = 52.62%. (**C**) Mean DDM parameters for two quantifiers conditions M and MTH. (**D**) The drift rate (v) for M and MTH against proportion (0 indicates 50%). The vertical dashed line indicates proportion 50% and the horizontal dashed line indicates drift rate zero.