Counterfactuals (not) under discussion

The meaning of counterfactual conditionals like (1) is a debated topic in semantics and philosophy. In recent work, Marty, Romoli & Santorio 2019 (MRS henceforth) tested the predictions of the major approaches in the literature by comparing simple cases like (1) to cases like (2), where counterfactuals are embedded in the scope of a negative quantifier. Their study, however, did not consider the potential role of (implicit) Questions under Discussion (QuD) on participants' responses, which is a crucial component for one of these approaches. Their results remain therefore inconclusive. We report on an experiment, built upon MRS's study, testing further the predictions of the different approaches by experimentally manipulating the QuD while at the same time extending the set of sentence types being tested. Our results provide further support for the selectional/supervaluational theories of counterfactuals (Stalnaker 1968, 1980).

(1) If ticket #37 had been bought, it would have won.

(2) None of these tickets would have won if it had been bought.

Two approaches. Two main approaches to counterfactuals are compatible with MRS' results. The first one is the HOMOGENEITY approach (H-approach; von Fintel 1997, Schlenker 2004 a.o.), according to which counterfactuals are universal quantifiers, with a definedness condition requiring that either all or no relevant worlds satisfy the consequent. The second is the SELECTIONAL approach (S-approach; Stalnaker 1968, 1980), which makes use of a selection function that takes a world and an antecedent as arguments and returns a 'selected' world: a counterfactual is true iff the consequent is true at that world. Since context is often insufficient to fix a value for the selection function, S-theories are accompanied by a supervaluational definition of determinate truth and falsity. The frame below illustrates the truth conditions for (1) predicted by each approach.

H-approach	$[\![(1)]\!]^{w,\preceq}$ =	DEFINED iff $\forall w'$: $w' \in \text{MAX}_{w,\preceq}([\![\text{\#37 is bought}]\!])$, #37 wins in w', or $\forall w'$: $w' \in \text{MAX}_{w,\preceq}([\![\text{\#37 is bought}]\!])$, #37 doesn't win in w'; TRUE iff $\forall w'$: $w' \in \text{MAX}_{w,\preceq}([\![\text{\#37 is bought}]\!])$, #37 wins in w'
S-approach	$[\![(1)]\!]^{w,s}$ =	TRUE iff $[\![\text{\#37 wins}]\!]^{s(w,[\![\text{\#37 is bought}]\!]),s}$
	(1) is DETERMINATELY TRUE [FALSE] at c iff, $\forall s$ compatible with c, $[\![(1)]\!]^{w,s}$ = true [false]	

Previous study. MRS tested the predictions of the two approaches with respect to undefinedness in complex sentences. Specifically, they compared cases like (1) to cases like (2) in contexts where tickets are selected by a random draw and only some tickets among those which were bought are winning, what they call a 'mixed lottery' scenario. As they discuss, the two approaches make distinct predictions with respect to (2): this sentence is predicted to be undefined by the H-approach (since the prejacent of *no ticket* is undefined, for each ticket in the domain of the quantifier), and simply false by the S-approach (since, for all choices of selection function, some tickets are going to win). Participants in MRS's study overwhelmingly rejected sentences like (2) in such contexts, in line with the S-approach and prima facie contra the H-approach.

A potential confound. The H-approach is generally supplemented with a pragmatic component whereby the undefined value can be interpreted as 'effectively true' or 'effectively false,' depending on the understood QuD (Križ 2015, Champollion et al. 2019). For instance, while semantically undefined, (2) can be judged as 'effectively false' in a context where the partition associated with the QuD does not distinguish between false and undefined cases, e.g., a context in which it is relevant whether each ticket bought *has a chance to win* – what we call an 'existential' QuD, in contrast to a context whether it is relevant whether each ticket *is guaranteed to win* – a 'universal' QuD. Since MRS did not control for potential QuD effects, participants may have understood (2) against an implicit existential QuD, which is a plausible QuD type given how lotteries generally work, and judged it as effectively false. This hypothesis allows the H-approach to account

for the high rates of rejection observed in MRS' study. Thus, when supplemented with the QuD ingredient, the H-approach remains compatible with MRS results.

Experiment. MRS's design and materials were modified in two significant ways. First, in addition to cases like (2), we tested the corresponding sentences with *every*, *not every* and *some*, as exemplified in (3). Second, we manipulated the type of QuD associated with each lottery scenario to control for the potential confound we described. The predictions that we tested are summarised in Table 1: in the target 'mixed lottery' scenarios where some randomly drawn tickets win, the H-approach predicts semantic undefinedness for all four sentence types exemplified in (3) as well as an effect of the QuD, which would push the undefinedness towards either true or false. The S-approach, on the other hand, predicts no QuD effect, but an effect of the quantifier strength: sentences with strong quantifiers like *every* and *no* are predicted to be false in the target scenarios whereas sentences with weak quantifiers like *some* and *not every* are predicted to be true.

Participants. 100 native speakers of English (mean age $= 41.94$ yrs, 50 female) took part in this experiment (payment: £3 for ≈15 min). The data of 3 participants failed to be recorded.

Task and Procedure. As the beginning of the study, participants were presented with a short background story describing the investment profile of one of two characters. During the test phase, participants were asked to read lottery scenarios and, after each scenario, they had to complete two tasks successively, as illustrated in Figure 1. First, they had to decide whether or not a person with the investment profile of the character previously introduced would invest in the lottery. Next, they were presented with one of the sentence types in (3) and had to assess the extent to which it is true in the given scenario using a continuous scale (slider) going from 'Completely false' to 'Completely true'. The experiment started with two unannounced practise trials. In the instructions, participants were encouraged to use the whole flexibility of the slider.

Materials. Participants were randomly assigned one of two QuD conditions, universal QuD (U-QuD) or existential QuD (EX-QuD). In the U-QuD condition, the character's goal was to win the lottery each and every single time, thus making relevant whether each ticket is *guaranteed to win*. In the EX-QuD condition, the character only wanted to have at least a shot at winning, thereby making relevant whether each ticket has *at least a chance to win*. The three lottery scenarios were the same as in MRS, namely the all and none control scenarios and the target mixed scenarios. Four quantifiers were tested so as to test the effect of quantifier strength in addition to polarity: *every* (strong, pos), *none* (strong, neg), *some* (weak, pos), and *not every* (weak, neg). Context (all, mixed, none), polarity (pos, neg) and strength (strong, weak) were manipulated within-participant. Each participant was presented with 12 test trials and 12 fillers.

Results. Responses in the QuD Check (Task 1) are shown in Figure 2. Participants responded as expected given the QuD conditions and lottery scenarios. The mean proportion of 'Yes' response for the All, Mixed, and None contexts was 100%, 90% and 2% respectively in the EX-QuD condition and 99%, 17% and 2% respectively in the U-QuD condition. Responses in the graded TVJT (Task 2) are shown in Figure 3. Responses were analysed using linear mixed-effects model testing the effects of QuD, polarity, strength, and their interactions (random effect: intercept for Subject with a by-subject slope for polarity). There was a significant effect of strength ($\beta = -77.12; p < .001$). No other effects was significant (all $|\beta|\text{s} < 5.0, ns$).

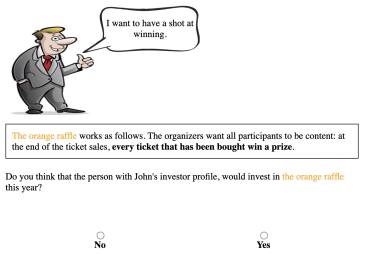
Discussion. We tested the predictions of two major approaches to counterfactuals, the S-approach and the H-approach. Building on MRS' study, we added a novel manipulation of the QuD to control for a potential confound in MRS and we further expanded the set of embedding environments. In our results, we found that participants were clearly sensitive to the QuD manipulation (Task 1), yet only a main effect of quantifier strength was found in the verification task (Task 2), as predicted by the S-approach. These results offer convincing evidence in support of the S-approach.

(3) a. Every one of these tickets would have won if it had been bought. POS-STRONG
b. None of these tickets would have won if it had been bought. NEG-STRONG
c. Not every one of these tickets would have won if it had been bought. NEG-WEAK
d. Some of these tickets would have won if they had been bought. POS-WEAK

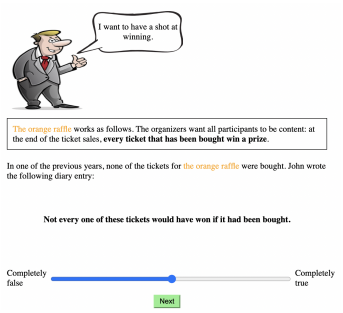
	POS		NEG	
	STRONG	WEAK	STRONG	WEAK
S-approach	**false**	**true**	**false**	**true**
H-approach	**undefined**	**undefined**	**undefined**	**undefined**
	EX-QUD	EX-QUD	U-QUD	U-QUD

Table 1: Predictions of the two approaches in the MIXED lottery scenarios. For the H-approach, the QUD type predicted to push the undefinedness towards true is indicated below each condition.

Figure 1: Example item illustrating the two-step response procedure in the study. This item is an instance of the existential QuD sentence in the *all* lottery context.



Task 1: QuD check



Task 2: Graded TVJT

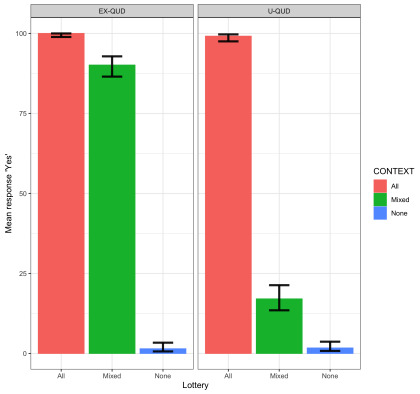


Figure 2: Mean 'Yes' response in Task 1. Error bars denote 95% CIs.

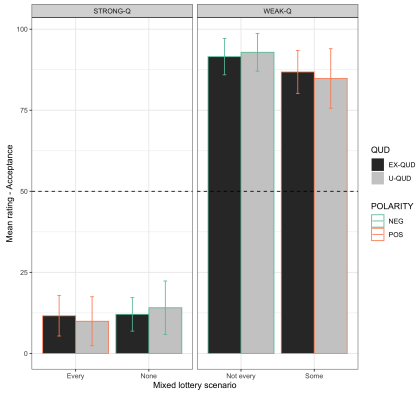


Figure 3: Mean ratings for the target sentences in the MIXED lottery scenarios by QUD type, POLARITY and STRENGTH in Task 2. Error bars denote 95% CIs.

Selected References

Champollion, L., Bumford, D., and Henderson, R. 2019. Donkeys under discussion • von Fintel, K. 1997. Bare Plurals, Bare Conditionals, and Only • Križ, M. 2015. Aspects of Homogeneity in the Semantics of Natural Language. • Marty, P., Romoli, J., and Santorio, P. 2019 Counterfactuals and Undefinedness: homogeneity vs supervaluations • Schlenker, P. 2004. Conditionals as Definite Descriptions • Stalnaker, R.C. 1968. A Theory of Conditionals. • Stalnaker, R.C. 1980. A Defense of Conditional Excluded Middle