A utility-based account of the choice of constituent question polarity

Seo-young Lee[1] **& Robert D. Hawkins**[1]
[1]Department of Linguistics, Stanford University
{sylee423,rdhawkins}@stanford.edu

Abstract

People choose their questions based on their conversational goals. But how do they decide between questions that would yield equivalent information? This study examines the mechanism by which a questioner chooses between constituent ("wh-") questions of positive vs. negative polarity, e.g., "Who came?" vs. "Who did not come?". Our Rational Speech Acts (RSA) model predicts the following: First, when a partial answer is expected, the question whose polarity aligns with the questioner's goal is preferred over its opposite polarity counterpart. Second, this goal alignment effect decreases when the structure of the questioner's decision problem requires an exhaustive answer. These predictions are confirmed in two experiments: Experiment 1 shows that goal alignment emerges in a free completion task, and Experiment 2 shows that it persists in a forced choice task, modulated by decision structure. Our utility-based account thus parsimoniously accounts for a phenomenon unexplained in formal semantic theory to date.

Keywords: constituent questions; question polarity; decision theory; Rational Speech Acts (RSA) framework

Introduction

Asking questions is one of our most powerful tools for learning about the world. From children exploring causal structure to adults navigating complex hypothesis spaces, people strategically select questions to gather information (Gureckis & Markant, 2012; Rothe et al., 2018; Ruggeri & Lombrozo, 2015). But the information a question elicits depends not just on what you ask about, but on *how you ask it*. A hiring manager asking "Which candidates are qualified?" versus "Which candidates are *not* qualified?" may end up with different shortlists. What drives the choice between these nearly identical forms?

Formal semanticists have long grappled with a version of this puzzle for polar questions. In standard Hamblin semantics (Hamblin, 1973), the denotation of a question is the set of its possible answers. So "Is John coming?" and "Is John not coming?" have identical denotations, since both admit {yes, no} as answers. Yet speakers are not indifferent between them (Büring & Gunlogson, 2000; Domaneschi et al., 2017; Ladd, 1981; Romero & Han, 2004, a.o.). Van Rooy and Šafářová (2003) provide a crucial insight: speakers prefer to pronounce the proposition whose truth has higher utility. Someone hoping John will come is more likely to ask "Is John coming?" not because it yields different answers, but because a "yes" answer directly confirms the desired state of affairs.

Constituent questions, also known as wh-questions, raise a parallel puzzle. "Who came to the party?" and "Who didn't come?" seem to request equivalent information: if the answer to one is exhaustive, its complement gives the answer to the other. Yet there is an intuitive difference: if your goal is to find someone to thank for bringing wine, you might ask who came; if your goal is to follow up with absentees, you might ask who didn't. Whether such goal-based preferences actually drive constituent question formulation, and what mechanism would explain them, has not been tested.

We propose that such asymmetry arises from the availability of the partial, or *mention-some* reading of constituent questions (Groenendijk & Stokhof, 1984, a.o.). The complement inference holds only when answers are assumed to be exhaustive. Under partial answers, it fails: if you ask "Who came to the party?" and hear "Alice and Bob," you cannot conclude that others didn't come, the respondent may simply not know. But if you ask "Who *didn't* come?" and hear "Carol," you can reliably identify an absentee. For a questioner seeking to follow up with someone who missed the party, the choice of question form is consequential. The relevance of mention-some interpretation, in turn, critically depends on the structure of the questioner's *decision problem*. When only one individual is relevant (say, finding one person to ask what happened) then a partial answer naming a single attendee suffices. When all individuals must be identified (say, sending thank-you notes to everyone who came) partial answers are inadequate and only exhaustive answers are resolving. Goal alignment should thus be strongest when mention-some answers are most useful.

We formalize this account using the Rational Speech Acts (RSA) framework (Degen, 2023; Goodman & Frank, 2016), extending prior work on how respondents tailor answers to questioners' decision problems (Hawkins et al., 2015, 2025). Our model generates two key predictions: (1) questioners should exhibit *goal alignment*, preferring to ask about the property they seek; (2) goal alignment should be strong under singleton selection but attenuated under set identification. We test these predictions in two experiments using a scenario where participants ask about contaminated vs. uncontaminated vials: Experiment 1 uses free completion to establish whether goal alignment emerges spontaneously; Experiment 2 uses forced choice to test modulation by decision structure.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

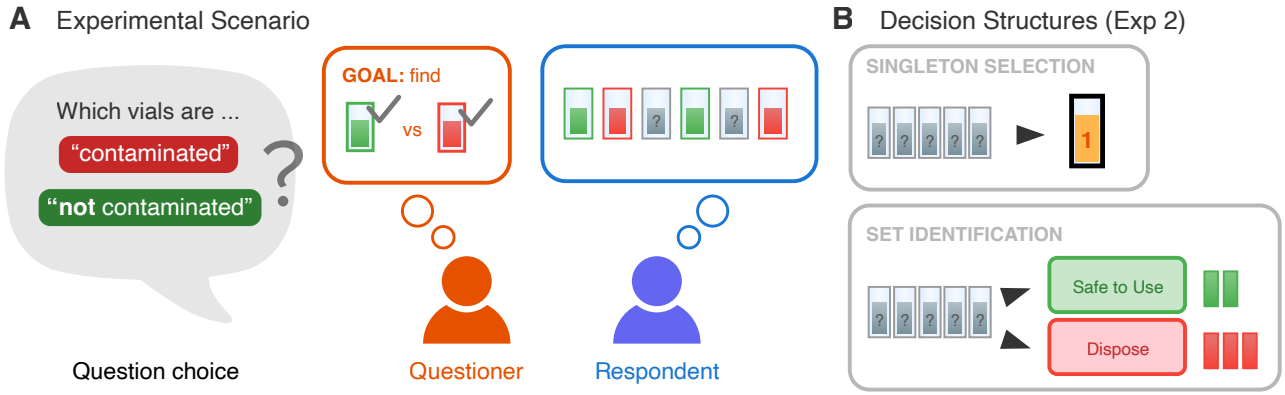

Figure 1: *Experimental design*. (A) Questioners choose between "Which vials are contaminated?" vs. "Which vials are not contaminated?" with the goal of finding either an uncontaminated or contaminated vial. (B) Decision structure manipulation (Exp 2): singleton selection requires identifying one goal-relevant vial; set identification requires sorting all vials.

Theoretical Framework

Background

A constituent question is a question whose propositional content contains a variable and whose answer is derived by substituting a particular value for the variable (Huddleston & Pullum, 2002). An answer to a question is a proposition, which, in intensional semantics, denotes the set of possible worlds where the proposition holds (Heim & Kratzer, 1998). If Sue and Bill are the only relevant individuals, the denotation of "Who came to the party?" in Hamblin (1973) semantics is:

$$\begin{aligned}\{\{&w : \text{nobody came in } w\},\\ &\{w : \text{only Sue came in } w\},\\ &\{w : \text{only Bill came in } w\},\\ &\{w : \text{Sue and Bill came in } w\}\}^{\downarrow}\end{aligned} \tag{1}$$

where $^{\downarrow}$ denotes downward closure under entailment (Farkas, 2020). The negative counterpart "Who did not come?" has the denotation:

$$\begin{aligned}\{\{&w : \text{there is nobody who didn't come in } w\},\\ &\{w : \text{only Sue didn't come in } w\},\\ &\{w : \text{only Bill didn't come in } w\},\\ &\{w : \text{Sue and Bill didn't come in } w\}\}^{\downarrow}\end{aligned} \tag{2}$$

These denotations are equivalent; each partitions the space of possible worlds in the same way. Under the assumption of exhaustivity, where answers pick out the maximal truthful element (Dayal, 2016), knowing the answer to one question determines the answer to the other: if "Sue came" is the exhaustive answer to (1), then "Bill didn't come" is the exhaustive answer to (2).

One key assumption behind the above generalization, as Dayal (2016) notes, is that the respondent correctly knows the status of all the individuals under consideration. This is not necessarily the case in real world conversations. If the respondent has partial knowledge, "Sue came" does not entail that Bill didn't; he may have come without the respondent knowing. Such a truthful yet non-maximal answer is called a partial or *mention-some* answer, as opposed to an exhaustive or *mention-all* answer. The two question forms thus differ in what can be reliably inferred from partial answers. A mention-some answer to "Who came?" identifies attendees but leaves non-attendees uncertain, while a mention-some answer to "Who didn't come?" identifies absentees but leaves attendees uncertain. To examine how a rational questioner would choose which form of question to ask under such uncertainty, we construct a probabilistic model as below.

Computational Model

We formalize our account using the Rational Speech Acts (RSA) framework. The key insight is that a questioner's choice of question form depends on the utility value of responses to that question, given their decision problem (Van Rooy, 2003).

Setup Consider a scenario with N vials, each either contaminated or uncontaminated. A world state $w \in \{0,1\}^N$ specifies the status of each vial, where $w_i = 1$ indicates vial i is uncontaminated. Each vial is independently contaminated with prior probability P_C. The questioner can ask one of two questions: "Which vials are contaminated?" (denoted q_+, positive polarity) or "Which vials are *not* contaminated?" (denoted q_-, negative polarity). A response $r \subseteq \{1,\ldots,N\}$ denotes a set of mentioned vials. Response r is a true answer to question q in world w if r is a subset of the vials satisfying the queried property. The questioner has a *goal* $g \in \{\textsc{find-uncontaminated}, \textsc{find-contaminated}\}$ and a *decision structure* $\mathcal{D} \in \{\textsc{singleton}, \textsc{set-id}\}$ determining how they need to act on the information received.

Intuition To see how an asymmetry between question forms emerges, consider $N = 3$ vials, where the respondent knows that vial 1 is contaminated, vial 2 is uncontaminated, and is uncertain about vial 3. A truthful (mention-some) answer to "Which are contaminated?" might be "Vial 1," which directly

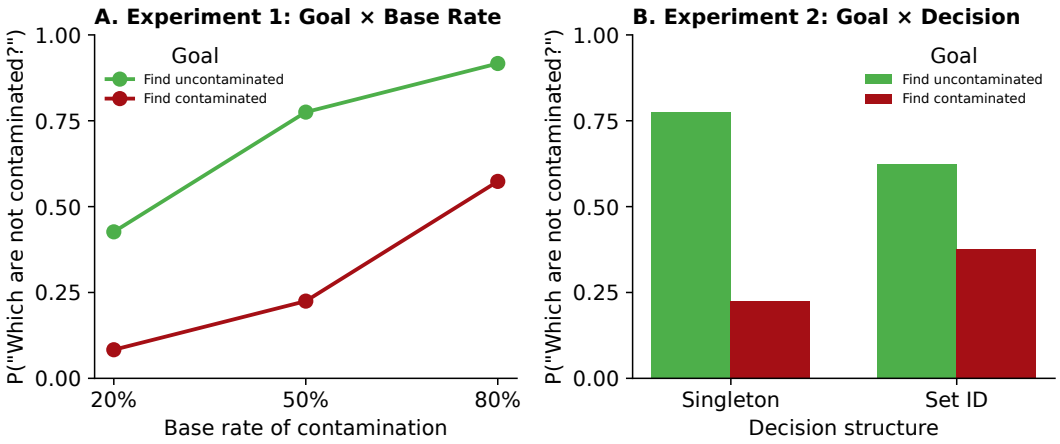

Figure 2: *Model predictions*. (A) The model predicts goal alignment across base rates, with increasing preference for "Which are not contaminated?" as contamination becomes more common. (B) The model predicts stronger goal alignment under singleton selection (55%) than set identification (23%). Parameters jointly fit to both experiments with all α=4.8, γ=0.92.

identifies a contaminated vial, but leaves the questioner unsure whether vial 2 or 3 is safe. A truthful answer to "Which are *not* contaminated?" might be "Vial 2," directly identifying a safe vial. Goal-aligned questions thus yield mention-some responses that directly identify goal-relevant entities, while goal-misaligned questions yield responses that confirm something irrelevant and leave the goal-relevant entities uncertain.

Respondent model We model the respondent as having *partial knowledge* about the world. Their knowledge state κ is such that for each vial $1, 2, \ldots, N$, they know that it is contaminated, that it is uncontaminated, or are uncertain about its contamination status. A knowledge configuration $c = (n_+, n_-)$ specifies that the respondent knows the status of n_+ contaminated vials and n_- uncontaminated vials.

The respondent's beliefs under κ are given as N independent Bernoulli distributions:

$$P_\kappa(w) = \prod_{i=1}^{N} p_i^{w_i}(1-p_i)^{1-w_i} \tag{3}$$

where $p_i = 1-\gamma$ for known contaminated vials, $p_i = \gamma$ for known uncontaminated vials, and $p_i = 1 - P_c$ otherwise. The parameter γ represents the respondent's confidence in their knowledge.

A literal listener L_0 interprets responses at face value, conditioning on the response being true:

$$L_0(w \mid q, r) \propto \mathbb{1}[r \text{ answers } q \text{ in } w] \cdot P(w) \tag{4}$$

where $P(w)$ is their prior over world states with each vial having an independent probability P_C of contamination. The respondent R_0 balances informativity, i.e., making the listener's posterior close to their own beliefs, and brevity. Formally:

$$R_0(r \mid q, \kappa) \propto \exp\left(\alpha \cdot [-KL(L_0(\cdot \mid q, r) \| P_\kappa) - \lambda|r|]\right) \tag{5}$$

where α controls response optimality and λ is an utterance length cost[1]. The respondent sometimes produces shorter, non-exhaustive (*mention-some*) responses. As vials within each knowledge category (known contaminated / known uncontaminated / unknown) are *exchangeable*, R_0 can be fully computed using the configuration $c = (n_+, n_-)$ and not the full knowledge state.

Questioner model The questioner does not know the respondent's knowledge configuration [2] and assumes a uniform prior. Given question q and response r, the questioner selects an action via soft-maximization:

$$\pi(a \mid g, q, r) \propto \exp\left(\alpha \cdot \mathbb{E}_{w \sim L_0(\cdot \mid q,r)}[U_{\mathcal{D}}(g, w, a)]\right) \tag{6}$$

The utility function $U_{\mathcal{D}}$ depends on decision structure. For *singleton selection*, the questioner selects one vial, receiving utility 1 if it matches the goal:

$$U_{\text{SINGLETON}}(g, w, a) = \begin{cases} w_a & \text{if } g = \text{FIND-UNCONTAMINATED} \\ 1 - w_a & \text{if } g = \text{FIND-CONTAMINATED} \end{cases} \tag{7}$$

For *set identification*, the questioner classifies all vials, receiving the F1 score for the goal-relevant category (precision and recall computed over uncontaminated vials for FIND-UNCONTAMINATED, over contaminated vials for FIND-CONTAMINATED).

The pragmatic questioner Q_1 selects a question to soft-maximize expected utility, marginalizing over knowledge con-

[1] λ was fixed at 0.1 in this study. Post-hoc sensitivity analysis confirmed that the AIC profile was essentially flat for $\lambda \in [0, 0.1]$; the mention-some asymmetry is primarily driven by the speaker's partial knowledge structure.

[2] Alternatively, their ignorance could be defined over knowledge states κ. Although our findings are under the assumption that the questioner assumes uniform prior over configurations c, the qualitative predictions are confirmed to be robust to this choice.

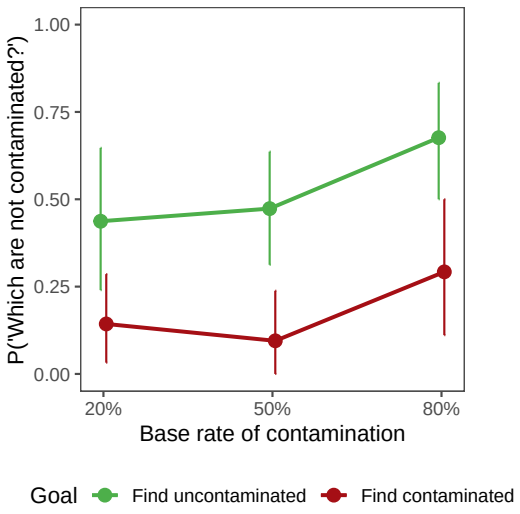

Figure 3: *Experiment 1 results*. Proportion of participants producing uncontaminated-synonymous completions (e.g., "safe," "clean") as a function of goal condition and base rate. Error bars indicate 95% CIs.

figurations and responses:

$$Q_1(q \mid g) \propto \exp\left(\alpha \cdot \mathbb{E}_c \mathbb{E}_{r \sim R_0(\cdot|q,c)}\left[V(g,q,r)\right]\right) \quad (8)$$

where $V(g,q,r) = \mathbb{E}_{w \sim L_0(\cdot|q,r)}[U_{\mathcal{D}}(g,w,a^*)]$ is the expected utility under the questioner's policy.

Predictions Figure 2 shows the model's two main predictions: (1) goal alignment across all base rates, with stronger preference for the negative question as contamination becomes more common, and (2) a 55-pp goal-alignment effect under singleton selection that is attenuated to 23 pp under set identification, reflecting the diminishing utility of mention-some answers when exhaustive responses are needed.

Experiment 1

We begin by testing the basic goal alignment prediction for constituent questions. By using a free completion task, where participants complete the sentence "Which vials are ___?", we can observe whether participants spontaneously produce goal-aligned predicates (e.g., "safe" vs. "contaminated"). We also vary the base rate of contamination to test whether the model correctly predicts that higher contamination rates increase the preference for asking about uncontaminated vials (the minority, more informative category)[3].

Methods

Participants We recruited 180 participants (30 per cell in the 2×3 design) via Prolific, located in the US, UK, or Canada whose first and primary language was English. One was excluded due to a technical issue, leaving 179 for analysis.

[3]All materials, data, and analysis code are available at `https://github.com/hawkrobe/wh-questions`.

Design The experiment used a 2 (Goal: find contaminated vs. find uncontaminated) × 3 (Base rate of contamination: $P_C \in \{20\%, 50\%, 80\%\}$) between-subjects design. Each participant was randomly assigned to one of the six conditions.

Materials and Procedure Participants were told they were a scientist working with a rack of 10 vials (Figure 1A), some of which were contaminated but visually indistinguishable from the uncontaminated vials. Rather than being explicitly given the percentage of contaminated vials, participants were told that "only a few / half / most" of the vials were contaminated for the 20% / 50% / 80% base rate conditions respectively. In the *find uncontaminated* condition, their task was to select one vial to use for an experiment; in the *find contaminated* condition, their task was to select one vial to examine and identify the contaminant. They could ask one question to a colleague who does not know the status of all the vials, but knows the status of some for certain. Three comprehension check questions were employed to ensure that participants understood their goal, the respondent's knowledge state, and the base rate condition. Data from participants who did not pass the comprehension checks were excluded.

On each trial, participants saw a visual representation of the vial rack, a proportion bar indicating the base rate of contamination (e.g., a bar with 20% green and 80% red in the 20% base rate condition), and a goal card that reminded them of their goal without explicitly mentioning the word 'contaminated' or 'uncontaminated'. They then completed the sentence "Which vials are _____?" by typing a word or a phrase. After the trial, participants were asked to report their strategy in deciding what question to ask.

Results

Responses were automatically coded as contaminated-synonymous (e.g., "unsafe" or "dirty") or uncontaminated-synonymous (e.g. "clean" or "safe") based on keyword matching. Of the 179 participants, 163 (91%) produced unambiguous responses; 16 were excluded as uncodeable (e.g., "red"). Because goal and base rate were both between-subjects with a single trial per participant, we analyzed the data using logistic regression (without random effects) predicting predicate meaning (1: uncontaminated-synonymous, 0: contaminated-synonymous) with goal (sum-coded), base rate (continuous, centered at 50%), and their interaction as predictors.

As predicted, we found a significant main effect of goal ($\beta = 0.83$, $z = 4.31$, $p < .001$). Participants in the find-uncontaminated condition were more likely to produce uncontaminated-synonymous completions (e.g., "safe," "clean," "not contaminated") than participants in the find-contaminated condition. Across base rates, 53% of find-uncontaminated participants asked about uncontaminated vials, compared to only 18% of find-contaminated participants. This goal alignment effect was consistent across all three base rate conditions (Figure 3).

We also found a significant main effect of base rate ($\beta = 1.69$, $z = 2.16$, $p = .031$): participants were more likely to ask

about uncontaminated vials when the contamination rate was higher. This is consistent with an efficiency account: when most vials are contaminated, asking about uncontaminated vials targets the smaller, more informative category. There was no significant goal × base rate interaction ($p = .97$), indicating that the goal alignment effect was robust across base rates. Comparing a null model (50% baseline) against a goal-matching heuristic confirms this effect (AIC = 226 vs. 197). These results establish that goal alignment emerges spontaneously in constituent question completions; Experiment 2 tests whether it is modulated by decision structure.

Experiment 2

Experiment 1 established goal alignment when participants freely formulate constituent questions. Experiment 2 tests whether the structure of their decision problem modulates the effect. By holding the predicate constant and varying only polarity, we isolate the choice between "Which vials are contaminated?" and "Which vials are not contaminated?"

Methods

Participants We recruited 160 participants via Prolific, located in the US, UK, or Canada, whose first and primary language was English. Participants were randomly assigned to one of four conditions (cell sizes ranged from 36 to 44 due to random assignment).

Design The experiment used a 2 (Goal: find contaminated vs. find uncontaminated) × 2 (Decision Structure: singleton selection vs. set identification) between-subjects design. Base rate was held constant at 50%.

Materials and Procedure The scenario was similar to Experiment 1, with two key changes. First, rather than completing a sentence, participants chose between two fully-formed questions: "Which vials are contaminated?" and "Which vials are not contaminated?" Second, we manipulated decision structure through the task framing. In the *singleton* conditions, participants were told they had to select a single vial—either to use for an experiment (find-uncontaminated) or to examine and identify the contaminant (find-contaminated). In the *set identification* conditions, participants were told they had to sort all vials into two bins labeled "Safe to Use" vs. "Dispose." To maintain goal alignment, participants in the find-uncontaminated condition learned that a colleague would use the "Safe to Use" vials for an important experiment, while participants in the find-contaminated condition learned that a safety inspector would examine the "Dispose" bin to verify proper handling. Each participant completed a single trial, followed by a strategy probe question.

Results

We analyzed the data using logistic regression predicting question polarity (1 = "Which are not contaminated?", 0 = "Which are contaminated?") with goal (sum-coded), decision structure (sum-coded), and their interaction as predictors. Replicating Experiment 1, we found a significant main effect of goal ($\beta = 0.88$, $z = 4.96$, $p < .001$): participants in the find-uncontaminated condition were more likely to ask "Which vials are not contaminated?" than participants in the find-contaminated condition. There was no main effect of decision structure ($p = .52$).

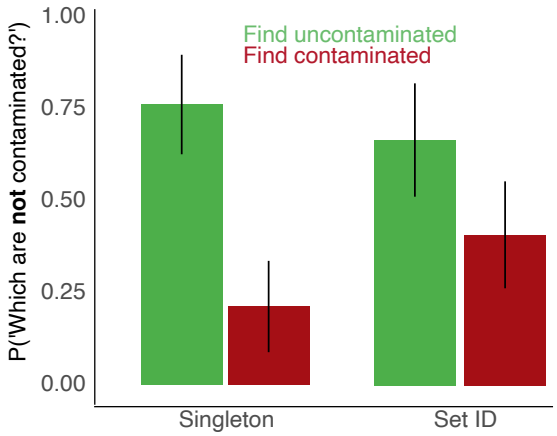

Figure 4: *Experiment 2 results*. Proportion of participants choosing "Which vials are not contaminated?" as a function of goal condition and decision structure. Error bars indicate 95% CIs. Goal alignment is attenuated in set identification relative to singleton selection.

Critically, we found a significant goal × decision structure interaction ($\beta = -0.35$, $z = -1.98$, $p = .048$). As shown in Figure 4, goal alignment was stronger in singleton selection than in set identification. In the singleton condition, 76% of find-uncontaminated participants chose the negative question compared to only 21% of find-contaminated participants—a 55 percentage point difference. In the set identification condition, this difference was attenuated to 26 percentage points (67% vs. 41%). While the study was powered primarily to detect the main effect of goal, the numerically substantial attenuation of goal alignment under set identification is consistent with our theoretical prediction: when exhaustive information is required, the two question forms become similar in their expected utility, reducing the asymmetry in their preference.

Model	k	AIC
Null	0	221.8
Goal-matching	2	199.3
Exhaustive-only RSA	2	199.5
Mention-some only RSA	2	199.5
Full RSA	**2**	**195.2**

Table 1: Experiment 2 model comparison.

Model Comparison To test whether the mention-some mechanism is necessary to explain the observed interaction, we compared the Full RSA model against several alternatives, fitting each model separately to the data (Table 1). The RSA models have two free parameters fit by maximum likelihood: a rationality parameter α and the respondent's confidence γ. The length cost λ, the number of vials N, and the base rate P_C were fixed rather than fit. A *goal-matching heuristic* estimates separate response rates for each goal, predicting goal alignment but no modulation by decision structure. The *exhaustive-only* baseline assumes the questioner always optimizes F1 score regardless of whether the actual task is singleton selection or set identification. Even when fit optimally, it predicts weak goal alignment in both conditions and cannot capture the interaction. The *mention-some only* baseline assumes the questioner only considers whether they identify one vial correctly, regardless of the actual task. It predicts strong goal alignment in both conditions and likewise cannot capture the interaction.

Critically, ablations of the RSA model achieve the same AIC (199.5) as the simpler goal-matching heuristic (199.3), indicating that the RSA machinery provides no benefit without correctly matching the utility-calculating mechanism to the task structure. Only the Full RSA model, which uses binary utility for singleton tasks and F1 score utility for set identification, captures the interaction. With optimized parameters ($\alpha = 4.8$, $\gamma = .92$), it predicts a 55% goal effect for singleton and 23% for set identification, closely matching the observed pattern (55% vs. 26%). The Full RSA achieves the best overall fit (AIC = 195.2), outperforming all alternatives by 4+ points.

General Discussion

Across two experiments, we found evidence that questioners' choice of constituent question polarity was driven by the expected utility of the answer under uncertainty about exhaustivity. Experiment 1 showed that goal alignment emerges spontaneously in free completion, with a stronger preference for the minority category whose partial answers are more informative. Experiment 2 replicated goal alignment in forced choice and showed that it is attenuated under set identification, where exhaustive responses are needed and the asymmetry between question forms diminishes.

We might wonder whether questioners simply "ask about what they want." Such a heuristic predicts goal alignment but not its modulation by decision structure, exactly the pattern only the Full RSA captures. The critical evidence is not goal alignment itself, but its attenuation under set identification. Moreover, negative questions may incur processing costs from the added complexity of negation, but processing costs alone would predict a uniform preference for positive forms, not the goal-dependent pattern we observe.

One notable finding is that goal alignment, while attenuated, was not eliminated under set identification (67% vs. 41%). One explanation for this persistence may be that participants do not fully internalize set identification as requiring exhaustive answers. The pragmatic expectation of partial responses is a strong default. Participants may treat the "Safe to Use / Dispose" framing as preserving the original goal asymmetry rather than fully equating the two question forms; the binary singleton utility may remain partially active even when an exhaustive task is described; or partial answers may simply be more cognitively accessible than F1-style classifications, biasing the comparison toward the singleton mechanism.

Several questions remain open. First, our experiments used a single scenario; generalizing across contexts is an important next step. Second, we did not directly measure participants' expectations about answer exhaustivity, which underlies the goal alignment effect. Manipulating these expectations and observing whether they modulate goal alignment would further strengthen our account. Finally, a complete theory must also account for epistemic bias (Romero, 2024)—the tendency to use negative polarity when one already believes the positive (e.g., "Who would do such a thing?") and vice versa—which our current framework does not address.

An alternative perspective treats mention-some readings as grammatically licensed rather than pragmatically derived. Xiang (2016, 2022), for example, argues that mention-some answers to wh-questions with a possibility modal (e.g., *Where can we get coffee?*) are constrained by structural features of the question that resist a purely pragmatic explanation. This grammatical view is largely complementary to ours: it concerns when mention-some interpretations are grammatically available, while we model how speakers choose between question forms given that they are. Our account simultaneously makes predictions for the listener side: if a questioner asks "Which are contaminated?", a pragmatic listener should infer that the questioner seeks contaminated vials. Testing whether listeners draw such inferences would provide converging evidence for the utility-based mechanism.

These results also point toward a more general theory of question polarity spanning constituent and polar questions. Van Rooy and Šafářová (2003) observe that polar-question speakers prefer to pronounce the proposition whose truth has higher utility (e.g. "Is the door unlocked?" when hoping to enter). The same principle, we suggest, underlies constituent question choice: questioners pronounce the property whose instantiation they seek. We suspect that a "no" response to a polar question leaves the same kind of residual uncertainty about the queried proposition that a partial constituent-question answer leaves about unmentioned individuals; a unified account is left for future work.

More broadly, utility-based decision theory illuminates aspects of language use that remain puzzling from a purely truth-conditional perspective. Positive and negative constituent questions have equivalent denotations, yet speakers systematically distinguish between them. By grounding question choice in the structure of downstream decisions, we suggest a division of labor between semantics and pragmatics that is both empirically supported and conceptually parsimonious.

Acknowledgments

We thank Greg Scontras, Cleo Condoravdi, members of the Social Interaction Lab, and audiences at the Stanford Construction of Meaning seminar for helpful discussion. We also thank Keny Chatain, the conversation with whom at SuB30 inspired this project.

References

Büring, D., & Gunlogson, C. (2000). Aren't positive and negative polar questions the same? [ms. UCSC.].

Dayal, V. (2016). *Questions*. Oxford University Press.

Degen, J. (2023). The rational speech act framework. *Annual Review of Linguistics*, *9*(1), 519–540.

Domaneschi, F., Romero, M., & Braun, B. (2017). Bias in polar questions: Evidence from English and German production experiments. *Glossa: a journal of general linguistics*, *2*(1), 1–28.

Farkas, D. (2020). Canonical and non canonical questions [ms. UCSC.].

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences*, *20*(11), 818–829.

Groenendijk, J. A. G., & Stokhof, M. J. B. (1984). *Studies on the semantics of questions and the pragmatics of answers* [Doctoral dissertation, Univ. Amsterdam].

Gureckis, T. M., & Markant, D. B. (2012). Self-directed learning: A cognitive and computational perspective. *Perspectives on Psychological Science*, *7*(5), 464–481.

Hamblin, C. L. (1973). Questions in Montague English. *Foundations of Language*, *10*(1), 41–53.

Hawkins, R. D., Stuhlmüller, A., Degen, J., & Goodman, N. D. (2015). Why do you ask? Good questions provoke informative answers. *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*.

Hawkins, R. D., Tsvilodub, P., Bergey, C. A., Goodman, N. D., & Franke, M. (2025). Relevant answers to polar questions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *380*(1932).

Heim, I., & Kratzer, A. (1998). *Semantics in generative grammar*. Blackwell.

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge University Press.

Ladd, D. R. (1981). A first look at the semantics and pragmatics of negative questions and tag questions. *Papers from the 17th Regional Meeting of the Chicago Linguistic Society*, 164–171.

Romero, M. (2024). Biased polar questions. *Annual Review of Linguistics*, *10*(1), 279–302.

Romero, M., & Han, C.-H. (2004). On negative yes/no questions. *Linguistics and Philosophy*, *27*(5), 609–658.

Rothe, A., Lake, B. M., & Gureckis, T. M. (2018). Do people ask good questions? *Computational Brain & Behavior*, *1*(1), 69–89.

Ruggeri, A., & Lombrozo, T. (2015). Children adapt their questions to achieve efficient search. *Cognition*, *143*, 203–216.

Van Rooy, R. (2003). Questioning to resolve decision problems. *Linguistics and Philosophy*, *26*(6), 727–763.

Van Rooy, R., & Šafářová, M. (2003). On polar questions. *Semantics and Linguistic Theory*, 292–309.

Xiang, Y. (2016). *Interpreting questions with non-exhaustive answers* [Doctoral dissertation, Harvard University].

Xiang, Y. (2022). Relativized exhaustivity: Mention-some and uniqueness. *Natural Language Semantics*.