Act or Clarify? Modeling Sensitivity to Uncertainty and Cost in Communication

Polina Tsvilodub[1], **Karl Mulligan**[1], **Todd Snider**[1], **Robert D. Hawkins**[2] **& Michael Franke**[1]
[1]Department of Linguistics, University of Tübingen (first.last@uni-tuebingen.de)
[2]Department of Linguistics, Stanford University (rdhawkins@stanford.edu)

Abstract

When deciding how to act under uncertainty, agents may choose to *act to reduce* uncertainty or they may *act despite* that uncertainty. In communicative settings, an important way of reducing uncertainty is by asking clarification questions (CQs). We predict that the decision to ask a CQ depends on both contextual uncertainty and the cost of alternative actions, and that these factors interact: uncertainty should matter most when acting incorrectly is costly. We formalize this interaction in a computational model based on *expected regret*: how much an agent stands to lose by acting now rather than with full information. We test these predictions in two experiments, one examining purely linguistic responses to questions and another extending to choices between clarification and non-linguistic action. Taken together, our results suggest a rational tradeoff: humans tend to seek clarification proportional to the risk of substantial loss when acting under uncertainty.

Keywords: clarification, *wh*-questions, directives, uncertainty, decision problems, regret

Introduction

We regularly face uncertainty when deciding how to act—whether about the state of the world, the consequences of potential actions, or another person's goals and intentions. When navigating such uncertainty, we face a fundamental choice: do we *act despite that uncertainty*, accepting the risk of error, or do we first *act to reduce the uncertainty*, actively seeking information before committing to action? Both strategies have costs: acting under uncertainty risks mistakes, while seeking information takes time and effort and delays other action. Which factors govern this tradeoff?

A rational agent should consider the expected value of what she could learn when choosing to seek (possibly costly) information before acting, as prescribed, for instance, by statistical decision theory when contemplating whether to perform an experiment (Raiffa & Schlaifer, 1961). In communicative settings, however, humans have access to a distinctively social form of information-seeking: asking *clarification questions* (CQs), to retrieve decision-relevant information from others (Dong et al., 2026; Purver et al., 2002; Saparina & Lapata, 2025). Still, the decision of whether or which CQ to ask may be subject to similar rationality considerations as the choice of experiments or of other kinds of questions (Benz, 2006; Coenen et al., 2019; Hawkins et al., 2015, 2025; van Rooij, 2003; Rothe et al., 2018). While many aspects of CQs have been investigated in prior work, such as the syntax of clarificational moves (Healey et al., 2003; Purver et al., 2002, 2003), or how conversational language models might implement them (Andukuri et al., 2024; Dagan et al., 2025; Kundu et al., 2020; Ma et al., 2025; Majumder et al., 2021; Testoni & Fernández, 2024), the question of *when* and *why* human interlocutors choose to ask a CQ has received almost no attention in empirical or computational work (Grand et al., 2026).

A prominent use of CQs is as a repair strategy for when an agent cannot fully resolve an utterance's content, helping to establish shared understanding (or *grounding*; Clark & Schaefer, 1989; Ginzburg, 1996, 2012). But CQs can also serve other functions (Schlangen, 2004), including resolving uncertainty about the context, such as when the current *decision problem* (DP) is not optimally addressable due to missing information (cf. Zhang et al., 2024). In such cases, rational choice theory makes two predictions: CQs should be more likely when (i) uncertainty about decision-relevant features is high, and (ii) available actions have low expected payoff. Crucially, these factors should interact: uncertainty matters most when mistakes are costly. We propose that this interaction arises because both factors are naturally unified through *expected regret* (equivalently: the *expected value of perfect information*; Raiffa & Schlaifer, 1961): agents ask CQs when their best available action risks substantial loss relative to what full information would afford. We test these predictions in a question-answering context where responses are purely linguistic (Experiment 1), then ask whether the same tradeoff governs choices between clarification and non-linguistic action (Experiment 2), before formalizing the account in a layered computational model based on expected regret.[1]

Experiments

In cooperative communicative contexts, we hypothesize that participants will trade off uncertainty about goal-relevant information against the cost of available actions when deciding whether to act despite uncertainty or to first ask a CQ. We focus on two types of speech acts which signal the speaker's decision problem (or discourse goal): questions and directives. We investigate the conditions which influence the propensity of interlocutors to ask a follow-up CQ, when there may be lingering uncertainty about the exact speaker goal. In both cases, for questions and directives, cooperative actions are available without clarification, although they may carry high

[1]All code and data are available at https://tinyurl.com/3z4awh4r.

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

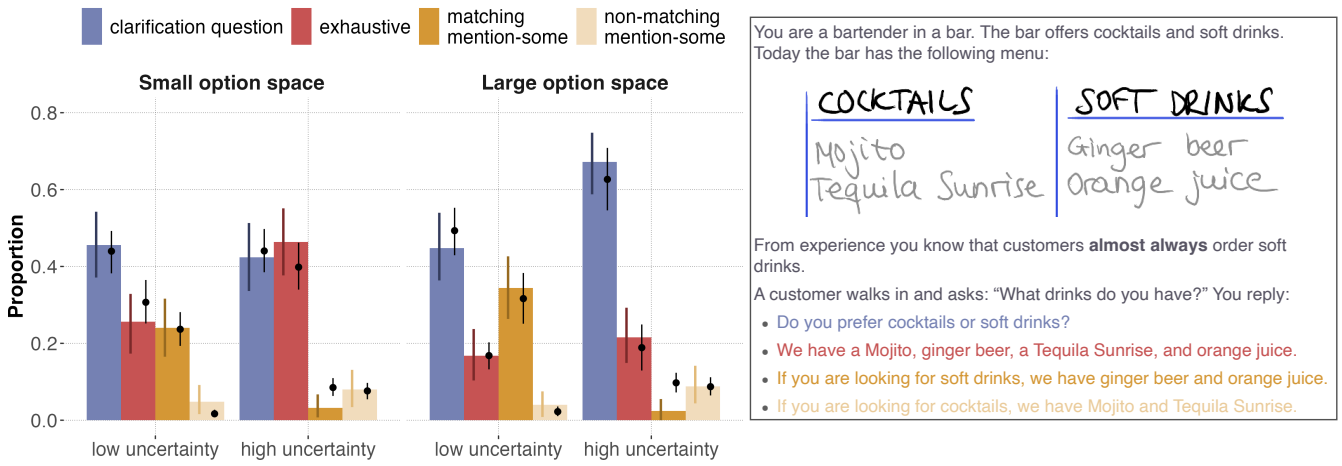

Figure 1: **Left**: Colored bars: Results of Experiment 1, showing proportions and 95% bootstrapped CIs of selected responses (colors) in different uncertainty conditions (x-axis) under different costs (facets). Under high uncertainty, no preference is mentioned, so the mention-some labels were assigned to responses randomly. Black dots: posterior predictive means (dots) and 95% credible intervals (black lines) of the computational model, under parameters fitted to human data via Bayesian data analysis. **Right**: Example item from Experiment 1 in the small option space, low uncertainty condition.

costs—e.g., giving an exhaustive answer to a question.

Experiment 1: Reacting to Questions

We begin with scenarios where someone asks a question (e.g., a customer walking into a bar, asking "What drinks do you have?") and an addressee aims to provide a helpful answer (e.g., the bartender offering options; see Figure 1). Previous work has shown that answerers reason about questioners' goals when selecting answers, even while implicitly accepting uncertainty about those goals (e.g., Hawkins et al., 2025; Stevens et al., 2016). Building on this work, we investigate when answerers opt to *reduce their uncertainty* by asking for clarification about the questioner's goal.

More specifically, the answerer has uncertainty about the practical *domain goal* that the question signals (e.g., getting a refreshing drink). Whatever the domain goal, the answerer always has the option of providing an *exhaustive* answer to the explicit question: a reliable but costly action. Choosing a more efficient action (e.g., a *mention-some* answer listing only non-alcoholic beverages) requires either information the answerer lacks or a willingness to act despite uncertainty about the questioner's practical goal. Alternatively, posing a CQ to determine the questioner's goal may offer a route to an efficient answer, but asking carries its own costs. If CQ choices involve reasoning about uncertainty and the costs of alternative answers, we make the following exploratory predictions:

1. TL;JustAsk: with a large option space, a higher rate of CQs under high uncertainty than under low uncertainty
2. NoNeedToAsk: with a small option space, no difference in rate of CQs between high and low uncertainty
3. JustListThemAll: with a small option space, a higher rate of exhaustive answers under high uncertainty than low
4. TooManyToList: with a large option space, no difference in exhaustive answers between high and low uncertainty

Methods & Materials To investigate these hypotheses, we ran a 2×2 factorial forced choice experiment, manipulating *uncertainty* about the questioner's goal and the cost providing an answer. Uncertainty was operationalized through a cover story about the questioners' likely preferences (e.g., for drinks); cost was operationalized through the number of contextually available options (*option space* size). Each trial listed specific options within two basic-level categories (e.g., cocktails and soft drinks), either two or four options per category (*small* or *large* option space). The cover story indicated that customers were equally likely to prefer either category (*high* uncertainty) or almost certainly preferred one category (*low* uncertainty; which category was counterbalanced).

Participants (N = 125) were recruited via Prolific, restricted to self-reported native English speakers in the US and UK, with approval rates over 95% and at least five prior studies. Each participant completed four trials (one per condition; within-subject) with items randomly sampled from a total of six. Participants selected among four response options (presented in randomized order). The experiment took approximately three minutes and participants were paid £0.45. An example trial is shown in Figure 1 (right).

Results Results are shown in Figure 1 (colored bars). Visual inspection supports TL;JustAsk and JustListThemAll. Additionally, when the option space is large but uncertainty is low, preference-matching mention-some answers are more common than under high uncertainty. To statistically test these qualitative patterns, we fit separate Bayesian logistic regression models for each answer type, with main effects of

uncertainty, option space, and their interaction, with maximal random effects structure. All contrasts were sum-coded.[2] We report posterior means and 95% credible intervals (CrIs excluding 0 indicate credible effects).

Confirming visual impressions, we found support for TL;JustAsk ($\beta = 1.58[0.75, 2.48]$) and JustListThemAll ($\beta = 2.00[0.87, 3.74]$), and, as predicted, no credible support for NoNeedToAsk ($\beta = -0.21[-0.99, 0.58]$) or TooManyToList ($\beta = 0.99[-0.32, 2.66]$). We also found credible main effects of uncertainty and option space size on the CQ rates ($\beta = 0.34[0.04, 0.67]$ and $\beta = 0.40[0.13, 0.67]$, respectively) and on exhaustive answer rates ($\beta = 0.75[0.24, 1.51]$, $\beta = -0.68[-1.24, -0.26]$). Finally, mention-some responses were more frequent under low uncertainty with a large option space ($\beta = -7.56[-18.80, -2.63]$) but not a small one ($\beta = -2.33[-6.82, 1.54]$). These results provide empirical support for our hypothesis that communicators are sensitive to uncertainty (asking for clarification when uncertainty is high), and that this propensity is modulated by costs of alternataive responses.

Experiment 2: Reacting to Directives

Experiment 1 provided evidence that people are sensitive to uncertainty, preferring CQs over other linguistic actions when costs are high. In Experiment 2, we broadened our scope to include decision problems addressable by *non-linguistic* actions. We focused on responses to directives. Both directives (e.g., "Please open the door.") and questions (e.g., "Would you please open the door?") can indicate the speaker's goals, but we focused on directives for reasons of pragmatic directness and syntactic simplicity. The respondent always had the option to select a direct action that might satisfy the directive, without clarification.

We also investigated whether this sensitivity is binarized or gradient by modulating the amount of uncertainty in the context. This distinction matters theoretically, as a binarized pattern would suggest a threshold-based decision process (asking a CQ whenever there is any uncertainty), while a gradient pattern would suggest sensitivity to the degree of uncertainty. To test if different *sources* of uncertainty were treated differently, we also varied whether the directive was underdetermined with respect to reference, manner, or deadline. Incorporating insights from Experiment 1, we also modulated the relative cost of acting—here, through the cost of mistakes—to see if this impacts how likely people are to ask clarification questions. We formulate the following exploratory predictions:

5. IfUnsure,Ask: binarized (i.e., non-graded) effect of uncertainty on CQs (by cost condition)
6. JustDoIt: with less uncertainty, more direct action
7. WorthAsking: with higher costs for acting incorrectly, more CQs (across uncertainty conditions)
8. Don'tMessUp!: with higher costs for acting incorrectly, less direct action

Methods & Materials To address these questions, we designed a 2×3 rating experiment, where we manipulated the amount of uncertainty in the action space (*no*, *low*, or *high* uncertainty), as well as the relative cost of selecting an incorrect action (*low* or *high* cost). Cost was manipulated through the consequences of an error. For example, opening the wrong cabinet is low cost while breaking down the wrong door with a sledgehammer is high cost. Participants ($N = 120$, 2 excluded for failing attention checks) read scenarios in which they were directed to do something, and then were asked to rate the likelihood that they would choose each of several (linguistic or non-linguistic) actions by using a slider bar ranging from "very unlikely" to "very likely". Both the participant instructions and scenario descriptions indicated that the directive-giver and the participant were aligned in their goals and in a social relationship such that issuing a directive would be natural, resulting in a shared decision problem for the director and respondent. In the *low* and *high* uncertainty conditions, the requested action (i.e., the way to resolve the shared decision problem) was underdetermined in terms of one of *reference* (uncertainty about what an expression refers to), *manner* (uncertainty about how an action is meant to be performed), or *deadline* (uncertainty about the timeframe in which the action is meant to be performed). Participants were presented a context and a set of options, as in the following referential example ({no | low | high} uncertainty, low cost):

(1) You are in a small kitchen with {a single | two | three} cabinets, all of which are currently closed. Your friend says, "Get me a pot from the cabinet!"
Action1/2/3: You open {the | the left/right | the left/middle/right} cabinet. CQ: You ask, "Which cabinet are the pots in?" NCQ: You ask, "Why should I do that?" Accept: You say, "OK" and do nothing else. Inaction: You do nothing.

To streamline participants' interpretation of the sliders in relation to one another, we presented a pie chart which dynamically updated as sliders were adjusted, representing the likelihood of each option; sliders were initialized at equal proportions, and each slider had to be clicked at least once before proceeding to the next trial. The provided ratings (0–100) were analyzed raw, to make responses across uncertainty conditions (with 1–3 direct action options) comparable.

Results

We found effects of both uncertainty and cost on reactions to directives, including our main tradeoff of interest between asking CQs and taking direct action. Results are in Figure 2.

For statistical analysis, we fit a Bayesian linear regression model, regressing the ratings against the main effects of uncertainty, reaction type, cost, their interactions and random

[2]The model in R syntax: `answer proportion ~ uncertainty * space size + (1 | itemName) + (1 + uncertainty + space size | subject)`. No interaction in the by-subject RE was included due to convergence issues.

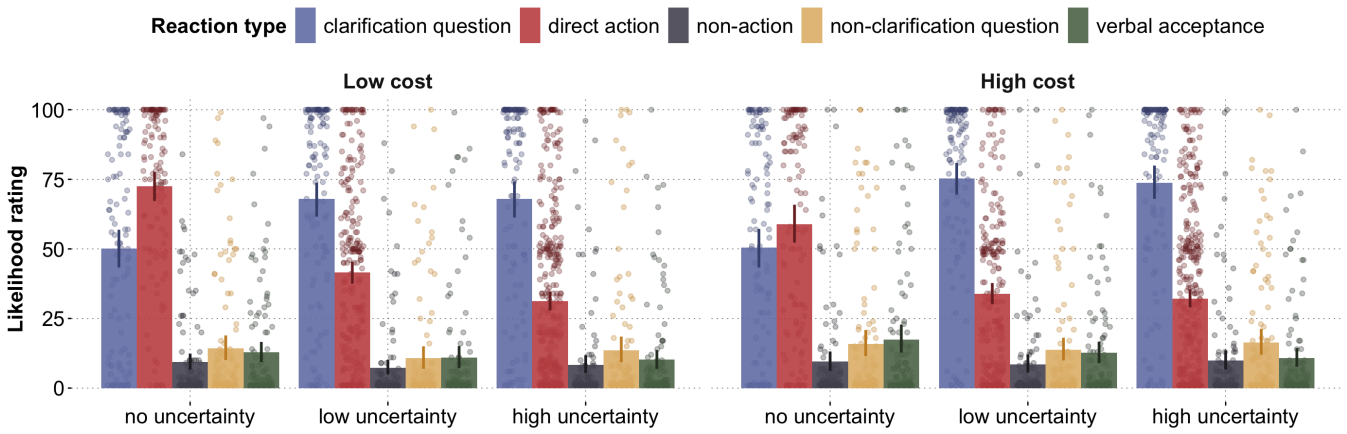

Figure 2: Results of Experiment 2: Unnormalized mean likelihood rating of different reaction options, grouped by uncertainty and costliness. Different direct actions were condensed into a single label for analysis. Error bars indicate 95% bootstrapped CIs, and points represent individual trials.

effects. All contrasts were sum-coded.[3] Starting with the effect of uncertainty, we found that across cost conditions, there was a credible effect of low uncertainty compared to no uncertainty on the rating of CQs ($\beta = 19.90[14.79, 24.91]$), but no credible difference in the rating in the high compared to low uncertainty condition ($\beta = 0.03[-4.99, 5.23]$), thus suggesting support for IfUnsure,Ask. On the other hand, there was support for a gradient interpretation of JustDoIt (high vs. low: $\beta = -7.93, [-11.38, -4.43]$, low vs. no:$\beta = -25.85, [-30.42, -21.32]$), confirming intuitions about hesitating to act under increasing uncertainty.

We also found evidence for IfUnsure,Ask both within the high cost condition (right facet of Figure 2; low vs. no: $\beta = 21.21[15.72, 26.59]$, but not high vs. low: $\beta = 0.68[-4.75, 6.25]$) and the low cost condition (left facet of Figure 2; low vs. no: $\beta = 18.60[13.25, 24.00]$, but not high vs. low: $\beta = -0.62[-6.18, 4.92]$). JustDoIt was likewise supported within each cost condition: in the high cost condition (high vs. low: $\beta = -7.28[-11.27, -3.29]$, low vs. no: $\beta = -24.55[-29.53, -19.60]$) and in the low cost condition (high vs. low: $\beta = -8.58[-12.60, -4.48]$, low vs. no: $\beta = -46.34[-55.98, -36.62]$).

Turning to cost, we also found smaller but credible effects. We confirmed WorthAsking: CQs were more likely when costs of acting incorrectly are higher ($\beta = 4.47[0.38, 8.48]$), and likewise, that direct actions become less likely, credibly supporting Don'tMessUp! ($\beta = -6.60[-9.69, -3.47]$). The remaining control reaction options of non-action, non-clarification question, and verbal acceptance were rated much less likely than CQ or direct action ($\beta = 42.69[35.78, 49.35]$). While these options were occasionally used, their limited use reflects the predicted infelicity of responding to a directive with asking a question that fails to bear on the decision problem (non-clarification question) or doing nothing (non-action), with or without acknowledgment of the task (verbal acceptance). We also found that the effect of uncertainty (low vs. no uncertainty) was credibly larger on CQs than on non-clarification questions ($\beta = 22.71[15.84, 29.50]$), indicating that the manipulation of uncertainty does not appear to affect question-asking in general; it is only consequential for questions which bear on the DP at hand, namely, CQs.

Lastly, we found that referential (*which*) CQs tend to have, on average, an appreciably lower likelihood rating than manner (*how*) or deadline (*when*) CQs. Exploratory by-type regression analyses suggest that the magnitude of the effects for CQs was driven by referential CQs, and qualitatively the same but marginally not credible for manner and deadline CQs; and while the gradient effects for direct actions were driven by deadline and manner contexts, a binarized effect was found in the referential contexts. This may be due to the typical domain sizes associated with each context type: while domains of entities are often finite and constrained, domains of manner or time can be much more indeterminate, such that the expected reduction of uncertainty from a CQ may be lower.

Computational Model

As predicted by rational choice theory, and confirmed by our experimental results, participants in our experiments seemed to be sensitive to both uncertainty in the decision space and to the costs of possible actions. The interaction of these factors can be captured using *expected regret*, which is equivalent to the *expected value of perfect information* (Raiffa & Schlaifer, 1961), and which reflects the expected difference between the expected utility *before* seeking information and the expected utility *after* eliminating all uncertainty. We formulate these ideas in a probabilistic model aimed to capture the empirical data from Experiment 1.

The decision problem is represented as a tuple $\langle G, P, R, U \rangle$,

[3] The model in R syntax: `rating ~ uncertainty * cost * reaction + (1 + uncertainty + cost || subject) + (1 + uncertainty + cost + reaction || item);` maximal converging REs were used. All direct action options were pooled.

where $G = \{g_1, g_2\}$ are the two contextual goals of the questioner, probability distribution $P \in \Delta(G)$ captures the decision-maker's uncertainty about G, the set of available actions $R = \{r_{exh}, r_{ms1}, r_{ms2}\}$ comprises all direct answers (see Figure 1, right), and the utility function $U : G \times R \to \mathbb{R}$ describes the agent's payoff. We assume that the four conditions of Experiment 1 differ in the degree of contextual uncertainty $\epsilon \in \{\epsilon_H, \epsilon_L\}$ for the high- and low-uncertainty conditions and in the cost associated with the exhaustive answer $\delta \in \{\delta_L, \delta_S\}$ for the large and small option spaces. The resulting parameterized decision problem thereby becomes:

Feature	$P(g_i)$	$U(r_{ms1})$	$U(r_{ms2})$	$U(r_{exh})$
g_1	$1-\epsilon$	1	0	1 - δ
g_2	ϵ	0	1	1 - δ

The core idea of the model is that, intuitively, an agent's reasoning is layered: faced with an uncertain decision context, an agent must first decide whether they have sufficient information to act. If none of the available actions are *good enough* (i.e., if expected regret is high; cf. Bourgeois-Gironde, 2010; Loomes and Sugden, 1982), the agent may opt to update the DP by seeking clarification, rather than simply taking one of the available options. Formally, the agent *reacts to* their contextual uncertainty, picking the clarification question r_{cq} with probability $P(r_{cq})$, or otherwise *acts under* uncertainty following a behavioral policy π:

$$P(r) = \begin{cases} P(r_{cq}) & \text{if } r = r_{cq} \\ (1 - P(r_{cq}))\ \pi(r) & \text{if } r \in R \end{cases}$$

Here, $\pi \in \Delta(R)$ is a soft-max policy with parameter $\alpha > 0$:

$$\pi(r) = \text{SoftMax}(\alpha\ \text{EU}(r)), \text{with}$$
$$\text{EU}(r) = \sum_{g \in G} P(g)\ \text{U}(g, r)$$

The probability $P(r_{cq})$ is defined as a monotonic function of the expected regret of the *best* action $r^* = \arg\max_r \pi(r)$ under the behavioral policy π. We use the logistic function $\text{Logistic}(x) = (1 + \exp(-x))^{-1}$:

$$P(r_{cq}) = \text{Logistic}(\tau \cdot (\text{ExpRegret}(r^*) - c))$$

where τ and c are parameters for shape and location. Finally, ExpRegret is defined as:

$$\text{ExpRegret}(r^*) = \sum_{g \in G} P(g) \cdot \text{Regret}(g, r^*)$$
$$\text{Regret}(g, r^*) = \arg\max_{r \in R} U(g, r) - U(g, r^*).$$

Bayesian Model Fitting For a given tuple of model parameters $\theta = \langle \epsilon, \delta, \alpha, \tau, c \rangle$, the model's prediction of answer choices is $p^\theta = \langle P(r_{cq}), P(r_{exh}), P(r_{ms1}), P(r_{ms2}) \rangle$, which we take to be the (multinomial) predictor for the data from one of the four conditions from Experiment 1. We use uninformative or flat priors: $\delta_{L,S} \sim \text{Uniform}(0, 1)$, $\epsilon_{L,H} \sim \text{Uniform}(0, 0.5)$, $\tau \sim \text{Uniform}(0, 5)$, $c \sim \text{Uniform}(0, 1)$, and $\alpha \sim \mathcal{N}(5, 1)$ (truncated to be positive). Without loss of generality, we assume that g_1 is the preferred goal in the low uncertainty condition, clamping $\epsilon \in [0, 0.5]$. We used Stan (Gelman et al., 2015) to obtain posterior parameter samples from four MCMC chains, each with 3000 warm-up and 4000 main samples. Robustness of the inference was assessed by inspecting the usual indicators ($\hat{R}$ was < 1.01, ESS above 2000).

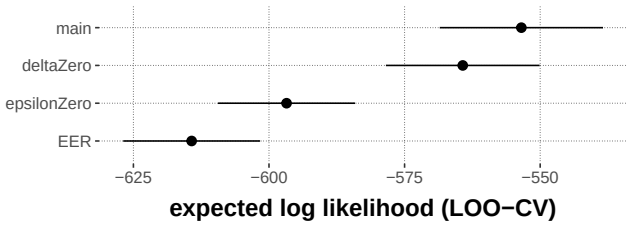

Figure 3: Estimates of expected log-likelihoods (dots) with standard errors (bars) for each model. Higher values are better.

The estimates for cost and uncertainty parameters are as expected for the four experimental conditions (large option space $\delta_L = 0.32[0.22, 0.43]$ was higher than $\delta_S = 0.11[0.00, 0.21]$ (small option space); high uncertainty $\epsilon_H = 0.49[0.46, 0.50]$ was higher than $\epsilon_L = 0.17[0.05, 0.30]$ (low uncertainty)), confirming that, without a priori constraints, the data informs the model as predicted. Estimates for the logistic link function parameters were $\tau = 3.60[1.95, 5.00]$, $c = 0.18[0.06, 0.30]$. The posterior means and 95% credible intervals of the posterior predictive distributions for each condition are shown in Figure 1 (black dots, CrIs). A visual posterior predictive check (Kruschke, 2010) suggests that the empirical proportions are within the credible interval of the proportions of CQs predicted by the model. This is corroborated by a non-significant Bayesian posterior predictive p-value ($Bpppv \approx 0.48$) obtained when using the (binomial) likelihood of CQ-choices as test statistic. However, Figure 1 also shows that the model does not match the human data perfectly. While the model predicts that there should be no preference for either mention-some answer in the high-uncertainty condition, participants seemed to have had a consistent bias towards choosing the option that was mentioned last in the description of the vignettes. Indeed, using the (multinomial) likelihood for the whole dataset we obtain $Bpppv \approx 0.003$, suggesting that the model does not fully capture the variance in the data.

To further investigate whether the model's main functional ingredients are plausible against the background of the observed data, we compare it to three alternative models. First, we consider an ablated model by setting $\delta = 0$ to isolate the contribution of reasoning about cost. Second, to investigate the role of uncertainty about G, we ablate by setting $\epsilon = 0$. Third, instead of using $\text{ExpRegret}(r^*)$, we use an expectation over the expected regret of all response options and pass it to a parameterized power law link function (EER model). Results from leave-one-out cross-validation are shown in Figure 3. A simple z-test suggests that the main model is indeed significantly better than the second best model with ablated δ ($p \approx 2.54 \cdot 10^{-8}$). We conclude that our model is reason-

ably piloting a computational account of how humans may decide between reacting to or acting under conversational uncertainty, suggesting that expected regret may be an adequate measure of how people confront the non-trivial trade-off between contextual uncertainty and action cost.

Discussion

Our work focused on answering the questions: when do agents act to reduce uncertainty—in particular, through communication, by asking CQs—and when do they act despite uncertainty, and what factors bear on that choice? Based on previous decision-theoretic work, we predicted that both the amount of decision-relevant uncertainty and the relative cost of available actions factor into that decision, and that they interact. That expectation was borne out in both experiments reported here: Experiment 1 focused on linguistic responses to questions, and showed that agents are sensitive to both uncertainty and cost, asking more CQs under higher uncertainty—especially when providing an exhaustive answer was costly. Experiment 2 focused on the competition between linguistic and non-linguistic reactions to directives, and again demonstrated the same trade-off: participants were more likely to ask CQs when uncertain or when errors were more costly.

Experiment 2 revealed two distinct patterns with respect to uncertainty. The rate of CQs showed a binarized pattern: credibly lower when there was no uncertainty, but no difference between non-zero uncertainty conditions, consistent with a threshold process where any uncertainty triggers consideration of clarification. The rate of direct actions, by contrast, showed a gradient pattern, declining with increased uncertainty across all three conditions, suggesting sensitivity to the degree of uncertainty when choosing among actions. This pattern is exactly in line with the layered structure of our computational model: considering both the uncertainty around and relative costs of the actions available—modeled via *expected regret*—, the agent first decides whether they have sufficient information for addressing the DP (in other words, whether to ask a CQ), and only afterwards would decide which (non-CQ) action to take. The former decision procedure uses a (soft) threshold, while the latter, ranking options by expected utility, uses a gradient measure.

Quantitative evaluation through Bayesian fitting of the model and comparisons to ablated model versions (without either uncertainty or cost) suggest that the expected regret-based model better explained the key interaction between uncertainty and cost in the human data. In other words, alternative models deriving the soft threshold for $P(r_{cq})$ from simpler uncertainty or cost based heuristics provided worse fit to human data. However, the proposed model might not fully capture conditions where the rate of exhaustive answers was higher than the CQ rate. It also did not account for biased rates of mention-some answers, suggesting that factors beyond expected regret such as response noise or recency effects may also influence answer selection.

The experiments and model presented here make certain assumptions which are worth highlighting. Experiment 1 made use of a special kind of conditional (of the form 'If ⟨goal⟩, ⟨mention-some⟩') to make clear the relationship between the response and the potential discourse goal (cf. Noh, 1998; Pittner, 2000). In theory, many more potential responses would be available to an agent (and relevant to the question) than were possible options in this experimental setup. A parallel experimental design using a free-choice production task might shed more light on which specific formulations speakers prefer. Similarly, Experiment 2 provided only one option for a non-clarificatory question (selected so as to remain on-topic without addressing the immediate DP); it is possible that speakers might have preferred other forms of non-clarificatory questions. That said, any such question would likely be interpreted in such a context as equally 'non-compliant', a de facto challenge to the directive, carrying with it social costs and thus reduced utility. Finally, our model's calculated expected regret depends on the scale of the utilities of the possible actions, for which we assume particular values (informed by Hawkins et al., 2025); alternative assumptions about those utilities or their scale would lead to a different range of expected regret, which might require different priors for the parameterization of the logistic link function.

As developed thus far, our model handles linguistic responses of the type available in Experiment 1, for which relative production costs are reasonably estimable and comparable (listing eight drinks is more effortful than listing four). We anticipate that the model should generalize to the setting of Experiment 2 as well, but this involves selecting among both linguistic and non-linguistic actions, which may have fundamentally different utility structures. For instance, the cost of giving an overly long answer differs in kind from the cost of breaking down the wrong door: one wastes time, while the other may be irreversible. There is no obvious common currency for comparing these costs, and the same action (e.g., asking a CQ) may carry different relative value depending on the alternatives it competes against. Adapting the model to mixed action spaces would require principled ways of specifying utilities across modalities; we leave this to future work.

Several further avenues remain for future work. We have focused on the costs of *not* asking (namely risking error from acting under certainty). But there are also costs *to* asking: CQs take time and effort, delay action, and may carry social costs such as appearing incompetent to a superior. While the parameter c in our model could be interpreted as the cost of asking, we did not explicitly manipulate this; future experiments might do so through time pressure or social context.

More broadly, our expected regret-based framework offers a unified approach to studying how agents navigate the tension between acting and inquiring. Clarification questions represent a distinctively human strategy for managing uncertainty, one that balances epistemic caution against the pressure to act. Understanding when and why people ask is a step toward understanding how humans coordinate action through language, and toward building systems that can do the same.

Acknowledgments

KM is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under project ID 547376651. TS and MF are funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under project ID 538184825 (CRC 1718). MF is a member of the Machine Learning Cluster of Excellence at University of Tübingen, EXC number 2064/2 – Project number 39072764.

References

Andukuri, C., Fränken, J.-P., Gerstenberg, T., & Goodman, N. (2024). STar-GATE: Teaching language models to ask clarifying questions. *First Conference on Language Modeling*. https://openreview.net/forum?id=CrzAj0kZjR

Benz, A. (2006). Utility and relevance of answers. In *Game theory and pragmatics* (pp. 195–219). Springer.

Bourgeois-Gironde, S. (2010). Regret and the rationality of choices. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*(1538), 249–257.

Clark, H. H., & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, *13*(2), 259–294.

Coenen, A., Nelson, J. D., & Gureckis, T. M. (2019). Asking the right questions about the psychology of human inquiry: Nine open challenges. *Psychonomic Bulletin & Review*, *26*(5), 1548–1587.

Dagan, G., Keller, F., & Lascarides, A. (2025). How^2: How to learn from procedural how-to questions. *arXiv preprint arXiv:2510.11144*.

Dong, Y. R., Hu, T., Hui, Z., Zhang, C., Vulić, I., Bobu, A., & Collier, N. (2026). Value of information: A framework for human-agent communication. *arXiv preprint arXiv:2601.06407*.

Gelman, A., Lee, D., & Guo, J. (2015). Stan: A probabilistic programming language for Bayesian inference and optimization. *Journal of Educational and Behavioral Statistics*, *40*(5), 530–543.

Ginzburg, J. (1996). Dynamics and the semantics of dialogue. *Logic, Language and Computation*, *1*, 221–237.

Ginzburg, J. (2012). *The interactive stance*. Oxford University Press.

Grand, G., Pepe, V., Tenenbaum, J. B., & Andreas, J. (2026). Shoot first, ask questions later? building rational agents that explore and act like people. *The Fourteenth International Conference on Learning Representations*. https://openreview.net/forum?id=EQhUvWH78U

Hawkins, R. X. D., Stuhlmuller, A., Degen, J., & Goodman, N. D. (2015). Why do you ask? good questions provoke informative answers. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *37*, 878–883.

Hawkins, R. X. 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).

Healey, P. G., Purver, M., King, J., Ginzburg, J., & Mills, G. J. (2003). Experimenting with clarification in dialogue. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*(25).

Kruschke, J. K. (2010). Bayesian data analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*(5), 658–676.

Kundu, S., Lin, Q., & Ng, H. T. (2020, July). Learning to identify follow-up questions in conversational question answering. In D. Jurafsky, J. Chai, N. Schluter, & J. Tetreault (Eds.), *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 959–968). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.90

Loomes, G., & Sugden, R. (1982). Regret theory: An alternative theory of rational choice under uncertainty. *The economic journal*, *92*(368), 805–824.

Ma, J., Shi, L., Robertsen, K. A., & Chi, P. (2025). Ambigchat: Interactive hierarchical clarification for ambiguous open-domain question answering. *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology*, 1–18.

Majumder, B. P., Rao, S., Galley, M., & McAuley, J. (2021, June). Ask what's missing and what's useful: Improving clarification question generation using global knowledge. In K. Toutanova, A. Rumshisky, L. Zettlemoyer, D. Hakkani-Tur, I. Beltagy, S. Bethard, R. Cotterell, T. Chakraborty, & Y. Zhou (Eds.), *Proceedings of the 2021 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 4300–4312). Association for Computational Linguistics. https://doi.org/10.18653/v1/2021.naacl-main.340

Noh, E.-J. (1998). A relevance-theoretic account of metarepresentatitive uses in conditionals. In V. Rouchota & A. H. Jucker (Eds.), *Current issues in relevance theory* (pp. 271–304). John Benjamins.

Pittner, K. (2000). Sprechaktbedingungen und bedingte Sprechakte: Pragmatische Konditionalsätze im Deutschen. *Linguistik Online*, *5*(1).

Purver, M., Ginzburg, J., & Healey, P. (2002). On the means for clarification in dialogue. In R. Smith & J. van Kuppevelt (Eds.), *Current and new directions in discourse & dialogue*. Kluwer Academic Publishers.

Purver, M., Healey, P., King, J., Ginzburg, J., & Mills, G. J. (2003). Answering clarification questions. *Proceedings of the Fourth SIGdial Workshop of Discourse and Dialogue*, 23–33.

Raiffa, H., & Schlaifer, R. (1961). *Applied statistical decision theory*. MIT Press.

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

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

Saparina, I., & Lapata, M. (2025). Reasoning about intent for ambiguous requests. *arXiv preprint arXiv:2511.10453*.

Schlangen, D. (2004). Causes and strategies for requesting clarification in dialogue. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue* (pp. 136–143).

Stevens, J. S., Benz, A., Reuße, S., & Klabunde, R. (2016). Pragmatic question answering: A game-theoretic approach. *Data & Knowledge Engineering*, *106*, 52–69. https://doi.org/https://doi.org/10.1016/j.datak.2016.06.002

Testoni, A., & Fernández, R. (2024). Asking the right question at the right time: Human and model uncertainty guidance to ask clarification questions. *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics*, 258–275.

Zhang, T., Qin, P., Deng, Y., Huang, C., Lei, W., Liu, J., Jin, D., Liang, H., & Chua, T.-S. (2024). Clamber: A benchmark of identifying and clarifying ambiguous information needs in large language models. *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 10746–10766.