Reconciling truthfulness and relevance as epistemic and decision-theoretic utility

Theodore R. Sumers¹, Mark K. Ho¹, Thomas L. Griffiths¹,², and Robert D. Hawkins³

¹Department of Computer Science, Princeton University
²Department of Psychology, Princeton University
³Princeton Neuroscience Institute, Princeton University

Author Note

We thank Xuechunzi Bai, Fred Callaway, Carlos Correa, Michael Franke, Adele Goldberg, and Noah Goodman for helpful discussions. TRS is supported by the NDSEG Fellowship Program and RDH is supported by the NSF (grant #1911835). This work was additionally supported by a John Templeton Foundation grant to TLG (grant #61454). We have no conflicts of interest to declare.

Experiment design and analyses were pre-registered at https://aspredicted.org/SPK_Z7Y and https://aspredicted.org/9MD_THB respectively. All materials, code, and data are available at https://github.com/tsumers/relevance. An early version of our model was presented at the 43rd Conference of the Cognitive Science Society. This manuscript is a draft (3/27/2023) and has not been peer-reviewed.

Correspondence should be addressed to Theodore Sumers. Email: sumers@princeton.edu.
Abstract

People use language to influence others’ beliefs and actions. Yet models of communication have diverged along these lines, formalizing the speaker’s objective in terms of either the listener’s beliefs or actions. We argue that this divergence lies at the root of a longstanding controversy over the Gricean maxims of truthfulness and relevance. We first bridge the divide by introducing a speaker model which considers both the listener’s beliefs (epistemic utility) and their actions (decision-theoretic utility). We show that formalizing truthfulness as an epistemic utility and relevance as a decision-theoretic utility reconciles the tension between them, readily explaining puzzles such as context-dependent standards of truthfulness. We then test a set of novel predictions generated by our combined model. We introduce a new signaling game which decouples utterances’ truthfulness and relevance, then use it to conduct a pair of experiments. Our first experiment demonstrates that participants jointly maximize epistemic and decision-theoretic utility, rather than either alone. Our second experiment shows that when the two conflict, participants make a graded tradeoff rather than prioritizing one over the other. These results demonstrate that human communication cannot be reduced to influencing beliefs or actions alone. Taken together, our work provides a new foundation for grounding rational communication not only in what we believe, but in what those beliefs lead us to do.

Keywords: Gricean maxims, rational models of communication, truthfulness, relevance, pragmatics, decision theory
Reconciling truthfulness and relevance as epistemic and decision-theoretic utility

Language allows us to influence others' beliefs and actions (Austin, 1962; Clark, 1996; Grice, 1975; Lewis, 1969). Yet this flexibility creates a challenge: given such sway over others, how should a cooperative speaker choose what to say? Paul Grice (1957, 1975, 1989) studied this decision and outlined an influential set of principles—now known as Gricean maxims—governing everyday discourse. Two of these principles are truthfulness (“Do not say what you believe to be false”) and relevance (“Be relevant”; Grice, 1975). While formalizing truthfulness is relatively straightforward, relevance continues to pose serious difficulties for models of communication. Indeed, immediately after stating the maxim, Grice admitted that it “conceals a number of problems” (Grice, 1975, p. 46) and avowed a desire to clarify it. Nearly fifty years later, consensus on a definition of relevance remains stubbornly elusive (Grice, 1989; Merin, 1999; P. Parikh, 1992; Roberts, 2012; van Rooij, 2003; Sperber & Wilson, 1987), fueling controversy over the nature and primacy of different maxims.

Following Grice’s seminal work, classic formal models defined both truthfulness and relevance in terms of the listener’s belief states. These models assume that speakers’ basic objective is to induce true beliefs about the world (Frank & Goodman, 2012; Goodman & Frank, 2016; Lewis, 1969; Stalnaker, 1978), then use relevance to prioritize specific information (van Kuppevelt, 1995; Roberts, 2012). For example, consider approaching someone on the street and asking them for the time. This notion of relevance captures the expectation that the person would answer your question, rather than telling you what they ate for breakfast—even though either response would be informative about the true state of the world.

But would the person give you the exact time? Evidence suggests they would not. People often provide rounded times (e.g., saying 5:00pm when it is actually 4:57pm; van der Henst et al., 2002). This is an instance of “loose talk,” a broad phenomenon in which people produce and accept approximations that are “true enough” for a given context (Lewis, 1979; Sperber & Wilson, 1985). Critically, what is “true enough” varies depending on what the listener intends to do with the information. For example, speakers are significantly more likely to provide an exact time when asked in order to set a watch (Gibbs Jr & Bryant, 2008; van der Henst et al., 2002) or...
when acting as a police witness (Mühlenbernd & Solt, 2022). Because standards of precision are context-dependent, Relevance Theory rejects the maxim of truthfulness and instead advances “positive cognitive effects” as the fundamental objective of communication (Sperber & Wilson, 1986; Wilson & Sperber, 2002b). However, it has proven difficult to formalize this construct (for commentary, see Levinson, 1989; Merin, 1999; Sperber & Wilson, 1987).

Our definition of relevance instead draws on an action-oriented view of communication tracing back to Austin (1962). We extend the scope of our model to reflect how the information acquired within discourse affects behaviors outside of it. This approach, prevalent in game-theoretic pragmatics (Benz, 2006; Franke, 2009; van Rooij, 2003), experimental semiotics (Galantucci & Garrod, 2011; Lazaridou & Baroni, 2020) and artificial intelligence (Allen & Perrault, 1980; Cohen & Perrault, 1979), frames communication as an instrumental tool allowing the speaker to act through others (Clark, 1996). Yet action-oriented models are typically scoped to specialized subsets of language like instructions or questions and are not usually advanced as general-purpose theories of human communication (but see Cohen & Levesque, 1988; P. Parikh, 2001; Vanderveken, 1990).

Thus, we reconcile truthfulness and relevance by integrating belief- and action-oriented approaches in a unified framework. Our computational model builds on the Rational Speech Act (RSA) framework (Frank & Goodman, 2012). The standard RSA model is belief-oriented, instantiating the maxims of truthfulness and relevance in an epistemic utility function (Goodman & Frank, 2016). We extend RSA by formalizing the listener as a rational agent who will act based on their beliefs. We then add a decision-theoretic utility function, quantifying relevance as the expected reward from the listener’s decision policy after hearing the utterance (Benz, 2006). Fig. 1 provides a schematic summary of these different models.

We first use this combined speaker model to explain classic examples from the literature. We show that the interaction between epistemic and decision-theoretic utility naturally focuses discourse (Roberts, 2012) while simultaneously predicting context-dependent deviations from truthfulness (van der Henst et al., 2002; Wilson & Sperber, 2002b). We then test novel

---

1 Other classic examples include statements like “Holland is flat,” which would be acceptable when planning a cycling trip yet inappropriate in a geological survey (Wilson & Sperber, 2002b).
Figure 1

A schematic summary of theories.

**A** Truthfulness

Truthful and useful

Relevance

True but misleading

False and misleading

**B**

- QUD
  - Roberts (2012)

- Relevance Theory
  - Sperber and Wilson (1986)

- Combined
  - Ours

**C**

Utterance $u$ → Listener

Beliefs $P(w | u)$ → Policy $\pi(a | w)$ → Action $a$

Speaker utilities

- Epistemic
  - Truthfulness
    - Stalnaker (1978)
    - Frank and Goodman (2012)
  - Relevance
    - Sperber and Wilson (1986)
    - Roberts (2012)
  - Politeness
    - Brown and Levinson (1987)
    - Yoon et al. (2020)
  - Persuasion
    - Mercier and Sperber (2011)
    - Barnett et al. (2022)

- Decision-theoretic
  - Relevance
    -earl (1992)
    - Benz and van Rooij (2007)
  - Commitments
    - Ostom et al. (1992)
    - Kanngiesser et al. (2017)
  - Norm enforcement
    - Baish et al. (2011)
    - Li and Tomasello (2021)

Note. (A) We consider two dimensions of utterances: their truthfulness and their relevance. (B) Shading approximates the preference over utterances suggested by different models. The “question under discussion” (QUD; Roberts, 2012) assumes the speaker is truthful and favors utterances that are relevant. Relevance Theory (Sperber & Wilson, 1986) drops the constraint on truthfulness and suggests speakers will only produce relevant utterances (the “presumption of optimal relevance”). Our Combined model predicts a graded tradeoff between truthfulness and relevance. This captures a general preference for truthfulness while allowing for contextually-dependent “loose talk.” (C) Schematic of speaker utilities. Utterances affect the listener’s beliefs about the world $P(w | u)$, which in turn determine their policy for choosing actions $\pi(a | w)$. Previous approaches describe speaker objectives in terms of beliefs or actions. Our Combined speaker (dark gray boxes) considers both, formalizing truthfulness and relevance as epistemic and decision-theoretic utilities respectively. Beyond Grice’s maxims, a range of communicative motives—including politeness, persuasion, and enforcement of norms or commitments—can be seen as utilities over beliefs or actions. We return to extensions incorporating these utilities in the General Discussion.
predictions made by our model. To do so, we first introduce signaling bandits: a generalization of traditional Lewis signaling games (Lewis, 1969) to multi-armed bandit settings (Sutton & Barto, 2018). We report a pair of studies using this paradigm. Our first experiment shows that, as predicted by our combined model, participants follow both truthfulness and relevance rather than either alone. Our second experiment pits these two objectives against each other. We find that neither dominates. Instead, participants demonstrate a graded tradeoff between the two: they are willing to endorse false statements with sufficient decision-theoretic utility. This challenges the basic assumption shared by existing models that either truthfulness or relevance is the primary goal of communication. Instead, our work suggests that integrated models considering both beliefs and actions are needed to capture the interplay between these two distinct objectives.

**Truthfulness, Relevance, and Speaker Goals**

We begin by reviewing in depth the challenges facing existing modeling approaches. We first consider two belief-oriented approaches, which assume the speaker’s goals are epistemic but differ in whether they take truthfulness or relevance to be primary. We then discuss an alternative action-oriented approach, which considers goals grounded in real-world behaviors. We will argue these perspectives each capture important aspects of truthfulness and relevance, but fail to provide a suitably general framework for reconciling them.

**Relevance to a “question under discussion”**

Classical formal models of communication focus on the listener’s beliefs. They assume providing truthful information is the primary objective, and introduce relevance to focus discourse on particular facets of the true state of affairs. This view frames cooperative communication as information transfer between speaker and listener (De Saussure, 1916; Lewis, 1969; Shannon, 1948). The canonical setup assumes a set or distribution of possible world states, and specifies the goal of communication as aligning on the true state (Stalnaker, 1978). Perhaps the purest basis for this view is the central principle of “accuracy dominance” in epistemic utility theory (Caie, 2013; Carr, 2017; Greaves & Wallace, 2006; Joyce, 1998). For example, Pettigrew (2016) argues that accuracy is “the only fundamental epistemic virtue: all other epistemic virtues derive their
goodness from their ability to promote accuracy.” This truth-centric objective lies at the core of recent cognitive models: for example, informative speakers in the Rational Speech Act (RSA) framework (Frank & Goodman, 2012; Goodman & Frank, 2016) typically aim to reduce the listener’s uncertainty over the true world state.

But truthfulness alone is clearly a starting point rather than the complete picture: communicating the full state of the world is not a practical conversational goal. Speakers cannot know or express every detail of the true world state, and listeners would not care to know them. Thus, Grice’s maxim of relevance is invoked to focus discourse on some particular facet of the world. Belief-based accounts have typically brought in relevance by assuming the conversation is intended to answer a particular “Question Under Discussion” (QUD; Benz & Jasinskaja, 2017; van Kuppevelt, 1995; Roberts, 2012). The QUD partitions the set of possible worlds, where each cell of the partition contains worlds consistent with one answer to the question. The speaker then seeks to communicate which of these cells contains the true world state, rather than the precise identity of the world state itself. It effectively coarse-grains the set of possible worlds, allowing the speaker and listener to ignore irrelevant details. RSA-based models have integrated the notion of a QUD (Goodman & Lassiter, 2015) to explain interpretation of nonliteral or vague language in various forms (Hawkins et al., 2015; Kao & Goodman, 2015; Kao, Wu, et al., 2014; Lassiter & Goodman, 2017; Yoon et al., 2020).

Notably, the basic mechanics of the QUD formalism place the maxim of truthfulness above the maxim of relevance: speakers are assumed to produce utterances that select the cell containing the true state. The assumption of truthfulness has proven impressively flexible, allowing listeners to derive a number of interesting implicatures from literally false utterances. For example, a listener hearing the hyperbolic statement “We waited years for a table” could make the statement truthful by constraining its meaning: in this case, inferring the speaker

[2] It may be objected that QUD-based speakers are not penalized for providing blatantly false information about irrelevant dimensions (Hoek, 2018), implicitly subordinating truthfulness about these dimensions. The important point is that the target cell under an epistemic QUD is determined solely in virtue of its relationship to some truthhood. Recent work outside the scope of our analysis has extended these techniques, allowing a speaker to be underinformative or deceptive in service of other conflicting non-epistemic goals, such as appearing competent (Yoon et al., 2018) or prosocial (Yoon et al., 2020), or being persuasive in service of a hidden agenda (Barnett et al., 2022).
intends to convey their annoyance about the event rather than literally describe its
duration (Kao, Wu, et al., 2014). Similarly, given a metaphor like “John is a shark,” a listener
may infer that John shares a relevant subset of characteristics with sharks—such as
aggressiveness—rather than literally being a member of that species (Kao, Bergen, et al., 2014).
In these cases, the background assumption of truthfulness with respect to something, together
with structured world knowledge, can be used to constrain meaning appropriately. However, the
underlying assumption that speakers attempt to communicate truthfully faces serious challenges.

Relevance as a “supermaxim”

An alternative proposal, Relevance Theory (Sperber & Wilson, 1986), posits that
relevance alone is sufficient to explain human communication. Relevance theorists discard the
maxim of truthfulness, holding that any tendency to say true things is an epiphenomenon
deriving from relevance (Wilson & Sperber, 2002b).

A key piece of evidence in favor of this perspective is “loose talk”: the fact that people
regularly produce approximate or even literally false statements—comments like “The lecture
starts at 5:00pm” or “Holland is flat”—which are accepted as “true enough” (Lewis, 1979) for
conversational purposes. Empirical evidence shows that manipulating the situational context and
exact question asked can yield highly variable rates of loose talk (Gibbs Jr & Bryant, 2008;
vander Henst et al., 2002; Mühlenbernd & Solt, 2022), undermining the idea that truthfulness is
a universal law of cooperative communication. Relevance theorists instead propose that speakers
choose utterances to maximize “positive cognitive effects,” described as “a worthwhile difference
to the individual’s representation of the world” (Wilson & Sperber, 2002a). However, it has
proven difficult to formalize these effects in purely cognitive terms, resulting in criticism that the
theory is fundamentally under-determined (Levinson, 1989; Sperber & Wilson, 1987).

How, then, might we quantify a “worthwhile difference to the individual’s representation of
the world”? We argue that worthwhile differences are best formalized in terms of decision making
outcomes. We propose that mutual world knowledge (Gibbs Jr, 1987; Sanford & Garrod, 1981;
Schank & Abelson, 1977)—most importantly, an understanding of conversational partners’
real-world goals and affordances—is integral to relevance. This is not theoretically controversial.
Belief-oriented theorists acknowledge that real-world circumstances ultimately drive relevance (Roberts, 2012) and even manipulate them to demonstrate contextual acceptance of loose talk (e.g., a generic query about the time, versus asking to set a watch; van der Henst et al., 2002). Similar real-world knowledge is required to understand examples used in the literature, from Grice’s description of a man stranded on a highway (Grice, 1975) to norms around lecture attendance (Wilson & Sperber, 2002b). Yet belief-oriented models are circumscribed to the information exchange within discourse and do not incorporate this real-world context. In the next section, we consider approaches that instead derive relevance from explicit models of a decision problem.

Relevance to a real-world decision problem

While traditional linguistics focuses on discourse itself, experimental semiotics (Galantucci & Garrod, 2011; Lazaridou & Baroni, 2020), classic AI research (Allen & Perrault, 1980; Cohen & Levesque, 1988; Litman & Allen, 1987, 1990), and game-theoretic pragmatics (Benz et al., 2005; Franke, 2009; P. Parikh, 2001) situate the communicative exchange within a larger context: the agents are assumed to have a set of real-world goals. Communication then serves as a means to these ends, allowing agents to act through others.

This instrumental view of communication is implicit in experimental semiotics (Galantucci & Garrod, 2011) and emergent language research (Lazaridou & Baroni, 2020). Studies in these paradigms typically place human or artificial agents in signaling games (Skyrms, 2010): tasks with common (Bard et al., 2020; Galantucci, 2005; Kang et al., 2020; Lazaridou et al., 2017; O’Connor, 2014; Steels, 2003; Vogel et al., 2013, inter alia) or mixed (Cao et al., 2018; Jaques et al., 2019) interests, which afford communication over a channel with no a priori semantics. The agents learn to communicate in order to maximize task rewards, and the resulting communication protocols are analyzed for theoretically important properties (Hockett, 1960) such as compositionality (e.g., Franke, 2016; Steinert-Threlkeld, 2020).

Similarly, a long line of mainstream AI research assumes speakers possess a set of real-world goals and communicate in order to fulfill them (Gmytrasiewicz & Doshi, 2005; Gmytrasiewicz & Durfee, 2001; Tambe, 1997). Task-oriented dialogue systems are the most
prominent example: these systems embed conversations within a planning module that reasons about external goals and available actions when interpreting or producing speech acts (Allen, 1983; Allen & Perrault, 1980; Cohen & Levesque, 1988; Cohen & Perrault, 1979; Perrault et al., 1978; Seneff & Polifroni, 2000; Young et al., 2013). Unfortunately, these systems’ strength is also their weakness: committing to a task-specific decision problem restricts their language understanding to that domain. Thus, while linguistic principles have been employed in AI research (Andreas & Klein, 2016; Dale & Reiter, 1995; Fried, Andreas, et al., 2018; Fried et al., 2021; Fried, Hu, et al., 2018; Golland et al., 2010; Monroe & Potts, 2013; Nie et al., 2020; Shen et al., 2019; Sumers, Ho, et al., 2021; Wang et al., 2016), these applications have not themselves been used to drive general theories of speech acts.

Unlike AI, game-theoretic pragmatics uses this instrumental view of communication to explain actual human discourse (P. Parikh, 2001). This leads naturally to decision-theoretic utility as a measure of relevance: utterances are relevant if they improve the expected utility of the listener’s real-world actions (Benz, 2006; Benz & Van Rooij, 2007; P. Parikh, 1991; 1992; R. Parikh, 1994). Decision-theoretic relevance offers two advantages over the set-theoretic “Question Under Discussion” approach (Roberts, 2012). First, it can be seen as deriving the QUD from the real-world decision context (Benz & Jasinskaja, 2017; Litman & Allen, 1987; Van Rooij, 2003), formalizing the idea that real-world “domain goals” lie at the root of “discourse goals” (Roberts, 2012). Second, the decision-theoretic formulation assigns each utterance a scalar preference. It is therefore strictly more expressive than the set-theoretic one, which expresses a binary preference over partitions (see Appendix A for a formal presentation of the QUD framework and its relationship to our proposed model).

To date, however, decision-theoretic relevance has been used primarily in the context of question asking (van Rooij, 2003) and answering (Benz, 2006; 2011; Benz & Van Rooij, 2007). We

---

3 Roberts noted that “we usually have goals in the real world, things we want to achieve quite apart from inquiry, *domain goals*. And our domain goals, in the form of deontic priorities, generally direct the type of inquiry which we conduct in conversation, the way we approach the question of how things are. We are, naturally, most likely to inquire first about those matters that directly concern the achievement of our domain goals” (Roberts, 2012, p. 7, emphasis in original).
instead instantiate this measure of relevance as a utility function in the Rational Speech Act framework (Frank & Goodman, 2012; Goodman & Frank, 2016). This casts decision-theoretic utility as a general objective in human communication and allows us to empirically determine its relationship with the maxim of truthfulness.\footnote{This decision-theoretic objective grounds speech acts into real-world behaviors, allowing belief- and action-based objectives to co-exist. In the next section, we detail this combined model.} This decision-theoretic objective grounds speech acts into real-world behaviors, allowing belief- and action-based objectives to co-exist. In the next section, we detail this combined model.

### A Framework to Reconcile Truthfulness and Decision-theoretic Relevance

Theorists from Grice (1975) onwards have recognized that human communication is guided by principles of truthfulness and relevance (but see Sperber & Wilson, 1986, 1987; Wilson & Sperber, 2002b), yet it has been challenging to incorporate both maxims in a sufficiently flexible formal model. Our approach builds on the Rational Speech Act (RSA) framework (Frank & Goodman, 2012; Goodman & Frank, 2016). We begin with the canonical RSA model optimizing for truthfulness: it proposes speakers choose informative utterances that yield true beliefs. We then extend RSA to account for actions by reformulating the listener as a rational actor who forms an action policy conditioned on their beliefs (Savage, 1954). Finally, we define a combined speaker model which chooses utterances that balance two independent utilities: belief-oriented truthfulness, which reflects the accuracy of the listener’s resulting beliefs, and action-oriented relevance, which reflects the decision-theoretic utility of the listener’s resulting actions (Benz, 2006). See Fig. 1C for a schematic of our model.

#### The Rational Speech Act framework

RSA (Frank & Goodman, 2012; Goodman & Frank, 2016) formalizes cooperative communication as recursive Bayesian inference. RSA models follow a standard structure, assuming that the lexical meaning of words are known to both speaker and listener. The pragmatic speaker is assumed to be cooperative and knowledgeable, and to choose utterances $u$ rationally to induce beliefs about the true world state $w$. Upon hearing an utterance, a pragmatic listener can then invert a model of the speaker to make stronger inferences about the state of the world.

\footnote{Following classical belief-oriented models, the Optimal Answer model (Benz, 2006) assumes the speaker is truthful.}
world $w$ based on their choice of utterance relative to alternatives.

Formally, speakers have access to some fixed set of utterances. They choose one proportional to a utility function $U(u, w)$, where $\beta_S$ is a soft-max parameter controlling speaker optimality:

$$P_S(u \mid w) \propto \exp\{ \beta_S \cdot U(u, w) \}. \quad (1)$$

The basic RSA speaker utility function emphasizes truthfulness: speakers choose utterances to maximize the probability the listener assigns to the true world state. Formally, the utility of an utterance is defined as the literal listener’s information gain about the true world state $w$ after hearing the utterance:

$$U(u \mid w) = \log P_{L0}(w \mid u). \quad (2)$$

Literal listeners are typically assumed to begin with a uniform prior over possible world states. $L_0$ then denotes their posterior beliefs after hearing the utterance. This requires the speaker to reason about the literal listener’s expected beliefs after hearing the utterance:

$$P_{L0}(w \mid u) \propto \delta_{[u](w)} P(w), \quad (3)$$

where $\delta_{[u](w)}$ represents the meaning of $u$, evaluating to one when utterance $u$ is true of $w$ and zero otherwise. $L_0$ is referred to as a literal listener because it interprets utterances strictly according to their lexical meanings. To formalize Gricean pragmatics, RSA then defines a pragmatic listener, $L_1$. This listener recursively embeds a speaker model (which in turn embeds a literal listener, $L_0$):

$$P_{L1}(w \mid u) \propto P_S(u \mid w)P(w) \quad (4)$$

Recursively reasoning about a speaker allows the listener to derive inferences beyond the literal semantics of an utterance. Much of the existing work within RSA studies pragmatic inference by modeling these $L_1$ listeners. Our work, however, focuses on the speaker: we ask what their objective function ought to be. Crucially, because a $L_1$ pragmatic listener embeds the speaker model, changing the speaker can have drastic effects on the listener’s inference—an effect we demonstrate after extending this model to a decision-theoretic framework.
Truthfulness as epistemic utility

The original RSA belief objective (Eq. 2) imposes a constraint that rules out false utterances, since a literal listener given a false utterance would ascribe zero probability to the true world. To allow the possibility of loose talk (van der Henst et al., 2002; Wilson & Sperber, 2002b), we replace this absolute interpretation of truthfulness with a scalar utility resembling the penalty on false utterances suggested by Franke (2009):

\[ U_{\text{Truthfulness}}(u \mid w) = \begin{cases} 1 & \text{if } \delta[u](w) = 1 \\ -1 & \text{if } \delta[u](w) = 0 \end{cases} \]  

(5)

Eq. 5 sets the epistemic utility of true utterances to 1 and false ones to -1. The speaker’s soft-max optimality then controls their degree of truthfulness: when \( \beta_S \approx 1 \), this models a preference for true utterances (Abeler et al., 2019; Franke, 2009; as \( \beta_S \to \infty \), this recovers a more typical RSA constraint to true utterances.\footnote{The exact scalar values used here do not affect qualitative model predictions. We used symmetric +1/-1 values for simplicity and interpretability, and return to this choice in the General Discussion.}

Relevance as decision-theoretic utility

To quantify the relevance of an utterance, we ground it into a decision problem. We formalize the decision problem by assuming the listener is a noisy-rational agent choosing from a set of possible actions \( A \). At a given point in time, a subset of those actions are available to the listener, which we refer to as a decision problem \( A \subseteq A \) (van Rooij, 2003). The utility of each action is defined by a reward function (Sutton & Barto, 2018), where the scalar reward value for an action is defined by the world state: \( R : A \times W \to \mathbb{R} \). Utterances inform the listener about the world state, and thus the payoffs associated with actions. Formally, the listener conditions their beliefs about the world state on the utterance (Eq. 3), then marginalizes over worlds to estimate the reward for taking an action \( a \):

\[ R_L(a, u) = \sum_{w \in W} R(a, w)P_L(w \mid u). \]  

(6)

\( R_L \) represents the listener’s posterior beliefs about their decision problem: it specifies the listener’s expected reward for action \( a \) after hearing utterance \( u \). We model the listener’s decision
policy $\pi_L$ as a softmax (Savage, 1954) over these beliefs. The listener chooses from actions $A$ according to their expected utility:

$$\pi_L(a \mid u, A) \propto \exp\{ \beta_L \cdot R_L(a, u) \},$$  

(7)

where $\beta_L$ is the listener's softmax optimality.

Utterances are then relevant if they induce beliefs that improve the listener's decision making. Formally, the relevance of an utterance is defined as the expected utility of the listener’s decision policy after hearing it:

$$U_{Relevance}(u \mid w, A) = \sum_{a \in A} \pi_L(a \mid u, A) R(a, w).$$  

(8)

This formulation unifies and generalizes prior work from game-theoretic pragmatics and RSA. It generalizes the game-theoretic Optimal Answer model of relevance (Benz, 2006, 2011; Benz & Stevens, 2018; Benz & Van Rooij, 2007), which can be recovered as the special case when the listener acts optimally ($\beta_L \to \infty$). It also generalizes previous action-oriented models in the RSA framework (Qing & Franke, 2015; Smith et al., 2013), which are the special case where the reward function is the classic Lewis (1969) payoff structure.

Formulating the listener as a rational actor provides a principled link from the speaker’s utterances to the listener’s behaviors (Fig. 1C). The speaker’s utterance first affects the listener’s beliefs about the world (Eq. 3). These beliefs determine the utility they ascribe to different actions (Eq. 6); the listener then chooses actions they believe are high reward (Eq. 7). Cooperative speakers choose utterances that maximize the true rewards of these actions (Eq. 8). Notably, this definition of relevance requires that the speaker knows the listener’s decision problem (i.e., has access to the set of possible actions $A$). We return to this assumption in the General Discussion and consider extensions including speaker uncertainty over the decision problem.

---

6 Eq. 8 reflects the expected utility of the listener’s decision policy after hearing the utterance, rather than the change in expected utility (van Rooij, 2003). However, due to the speaker’s softmax decision policy (Eq. 1) these formulations are equivalent: subtracting a constant does not affect the speaker’s choice of utterance.
Reconciling Truthfulness and Relevance

The formalisms above disentangle listeners’ beliefs about the world, $P_L(w \mid u)$ and their subsequent actions $\pi_L(a \mid u, A)$ while preserving a principled link between the two. We then propose a combined speaker model which ascribes utility to both beliefs and actions. Truthfulness is quantified as epistemic accuracy (Eq. 5) and relevance as decision-theoretic utility (Eq. 8). The speaker’s utility function is a convex combination of the two, plus a cost term $C(u)$:

$$U_{\text{Combined}}(u \mid w, A) = \lambda \cdot U_{\text{Relevance}} + (1 - \lambda) \cdot U_{\text{Truthfulness}} + C(u).$$

(9)

The cost term is a standard component of RSA (Goodman & Frank, 2016) allowing for variable production or processing effort. We assume uniform costs over utterances unless otherwise noted.

The $\lambda$ parameter has an intuitive interpretation: it determines the relative weight placed on truthfulness and decision-theoretic objectives. When $\lambda \rightarrow 1$, speakers choose utterances solely to maximize the expected rewards from the listener’s behavior. Such a speaker will readily produce false utterances to induce desirable behaviors. When $\lambda \rightarrow 0$, we recover a purely informative speaker that chooses utterances solely based on their epistemic value. Intermediate values of $\lambda$ will cause the speaker to blend truthfulness and relevance, choosing utterances that are both truthful and likely to yield good decisions. We next demonstrate how this framework explains puzzles from the literature.

Explaining Phenomena via Decision-Theoretic Utility

A satisfactory account of the interplay between truthfulness and relevance must explain at least three distinct phenomena. First, following Grice, assumptions of relevance should allow listeners to use real-world context to enrich the literal content of an utterance (Grice, 1975). Second, relevance should determine context-dependent standards of truthfulness. This means that under certain circumstances, incorporating relevance should have the opposite effect and instead relax the literal content of an utterance (Wilson & Sperber, 2002b). Finally, the framework must account for “loose talk”: it should capture when and why a speaker might produce a false utterance that is intended to be taken literally (van der Henst et al., 2002; Wilson & Sperber, 2002b). In the following sections, we use a series of classic examples from the literature to
Table 1

*Formalizing relevance from Grice (1957).*

<table>
<thead>
<tr>
<th>Garage State $w$</th>
<th>None</th>
<th>Closed</th>
<th>Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Probability $P(w)$</td>
<td>.8</td>
<td>.1</td>
<td>.1</td>
</tr>
<tr>
<td>Reward $R(a, w)$</td>
<td>Leave car</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>Stay at car</td>
<td>-.5</td>
<td>-.5</td>
</tr>
</tbody>
</table>

illustrate how our synthesis of epistemic and decision-theoretic utility satisfies these desiderata.

**Decision-theoretic relevance can strengthen pragmatic inferences**

Our first example shows how formalizing decision-theoretic relevance can strengthen pragmatic inference. To motivate the maxim of relevance, Grice (1975) described the following real-world scenario: A is standing by an immobilized car; B approaches, and the following exchange occurs:

A (Listener): I am out of petrol.

B (Speaker): There is a garage round the corner.

Grice justifies this brevity with the maxim of relevance. He suggests that B’s statement implies the garage is open. Note that the statement literally communicates only the garage’s existence: it is equally consistent with the garage being open or closed. Yet intuitively, we expect that the speaker is trying to help the listener. If the speaker believed the garage was closed, it would not help the listener, and so they would not mention its existence. Thus, a pragmatic listener—given such a vague utterance—should infer that the garage is likely to be open. We now show how decision-theoretic relevance formalizes this intuition.

We first define the world states $w$ and payoff matrix $R(a, w)$ shown in Table 1. We assume that a priori it is relatively unlikely that a garage exists around the corner, but given its existence, it is equally likely to be open or closed. We then assume A has two available actions: to stay with their car, or leave it in search of gas. We assign payoffs to each possible world-action

---

7 See Benz (2006) for a game-theoretic analysis and Thomason (1990) for a plan-based analysis of this example.
Figure 2

*Relevance can strengthen pragmatic inference.*

Note. After hearing the utterance “There’s a garage round the corner,” different assumptions about the speaker’s objectives yield different posteriors over the world state. Left: a pragmatic listener assuming a purely truthful speaker draws no additional inference beyond the literal interpretation, contradicting human intuition. Right: assuming a decision-theoretic speaker yields the intuitive inference that the garage is open. This asymmetry is driven by the assumption that the speaker mentions the garage because they believe it can resolve the listener’s current predicament. See Fig. [B1](#) for a sensitivity analysis exploring different parameter settings.

pair. A successful search for gas is the most desirable outcome (reward of 1); staying with the car is undesirable (reward of -.5), but better than an unsuccessful search for gas (which would leave the car abandoned; reward of -1). A listener with no information thus prefers to stay with their car, as a successful search for gas is unlikely.

In order to model the interaction, we assume the speaker could produce four distinct utterances \( u \), each corresponding to a possible knowledge state: one uninformative (“Sorry”); one vague (“There is a garage round the corner”); and two precise (“...and it is open/closed”). These correspond to the speaker’s possible knowledge states. To compare the effects of pragmatic inference with and without decision-theoretic utility, we define two speakers: a “Truth-only” speaker and a “Combined” speaker. We set the “Truth-only” \( \lambda = 0 \), the “Combined \( \lambda = .5 \), and \( \beta_S = \beta_L = 10 \) for both.

To explore the effects of the speaker’s objective on pragmatic inference, we define a pair of

---

8 Across all examples, we choose parameter settings that yield intuitive effect sizes. However, the qualitative effects of interest hold across a range of parameter space; see Appendix [B](#) for sensitivity analyses.
pragmatic listeners (Eq. 4) each embedding one of these speakers. Finally, we analyze the
listeners’ posteriors over the world state $w$ after hearing the vague utterance “There is a garage
round the corner.” Fig. 2 shows the effect of the different speaker models on the listener’s
posterior. If the listener assumes the speaker is purely guided by epistemic truthfulness, there is
no reason to favor either the open or closed states. However, if the listener assumes the speaker is
trying to provide information to resolve their underlying predicament, they can infer the garage is
likely open.

**Decision-theoretic relevance can relax pragmatic inferences**

Our previous example demonstrates that decision-theoretic relevance can strengthen
pragmatic inference. However, proponents of Relevance Theory (Sperber & Wilson, 1986; Wilson
& Sperber, 2002b) have noted that expectations of relevance can also have the opposite effect:
listeners may take a precise statement and loosen its meaning. For example, a statement such as
“The lecture starts at five o’clock” is generally taken to mean “starts at five o’clock or shortly
after” (Wilson & Sperber, 2002a, p. 596, emphasis ours). We again formalize the listener’s
decision context and show how expectations of decision-theoretic relevance can drive such
loosening of meaning.

---

Table 2

Formalizing relevance from Wilson and Sperber (2002a).

<table>
<thead>
<tr>
<th>Lecture Start $w$</th>
<th>5:00</th>
<th>5:05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prior Probability $P(w)$</td>
<td>.5</td>
<td>.5</td>
</tr>
<tr>
<td>Reward $R(a, w)$</td>
<td>Arrive 5:00 1 .9</td>
<td>Arrive 5:05 -1 1</td>
</tr>
</tbody>
</table>

---

9 Wilson and Sperber (2002a) use this example as part of a larger point that start and end times yield different interpretations. We address this broader point in our following example, so for simplicity we analyze only the start time; however, the same logic applies to the end time. Wilson and Sperber (2002a) also consider more complex utterances and variable costs, but these are not needed to derive the basic effects of interest here. Finally, see P. Parikh (1992) for a game-theoretic analysis based on this example.
As noted by Wilson and Sperber (2002a), a key insight is that the costs of arriving early versus late are asymmetric: arriving a few minutes early is preferable to arriving a few minutes late. Again, we formalize this with a payoff matrix (shown in Table 2): arriving whenever the lecture starts is optimal (with a reward of 1); arriving early is nearly as good (reward of .9) while arriving late is costly (reward of -1). We consider a pair of utterances: “The lecture starts at 5:00pm” or “The lecture starts at 5:05pm.” As before, we define a pair of speaker models—a “Truth-only” speaker ($\lambda = 0$) and a “Combined” speaker concerned primarily with relevance ($\lambda = .9$), with $\beta_S = \beta_L = 5$ for both. We assume a “precise” literal semantics of number, such that an utterance is true only when it exactly matches the true value.\footnote{This assumption oversimplifies a vast literature grappling with the complexity of number semantics. We do not intend this as a standalone account; for example, earlier RSA models (Kao, Wu, et al., 2014) handle numerical imprecision by introducing uncertainty over whether one’s partner uses a ‘precise’ or ‘fuzzy’ literal semantics for number and derives rounded interpretations as a cost implicature, consistent with Krifka (2007). This account clearly distinguishes between mere imprecision (where rounding is still literally true under a “fuzzy” semantics) and outright lying (where giving a time that is hours off is true under no number semantics). Because we are interested in the context-sensitivity of imprecision with respect to the decision problem, our analysis goes through under a more sophisticated treatment.} Again, we consider a pair of pragmatic listeners (Eq. 4) each embedding one each of these speakers, and analyze their beliefs after hearing the utterance “The lecture starts at 5:00pm.”

Fig. 3 shows the resulting pragmatic listener posteriors over $w$, and Fig. B2 gives a parameter sensitivity analysis. In this case, because the speaker’s utterance is literally precise (it specifies an exact world state), a pragmatic listener embedding a “Truth-only” speaker assumes the literal interpretation is correct and the lecture is certain to start at 5:00pm. However, expectations of decision-theoretic relevance can loosen the literal meaning. In particular, because arriving early is nearly as good as arriving precisely on time, a speaker believing the lecture is likely to start at 5:05pm may still say it starts at 5:00pm. A pragmatic listener embedding such a speaker derives the colloquial interpretation “starts at 5:00pm or shortly after.” Formalizing the asymmetric decision-theoretic utility associated with early versus late arrival explains why such loose language may be “true enough” (Lewis, 1979) in such a context.
Figure 3

Relevance can loosen literal meanings.

![Graph showing Relevance can loosen literal meanings.]

Note. Pragmatic inference after hearing “The lecture starts at 5:00pm” using different speaker models. Left: pragmatic inference assuming a purely truthful speaker results in certainty that the lecture will start at 5:00pm. Right: pragmatic inference assuming a “Combined” speaker yields the intuitive inference that the lecture may actually start at 5:05pm. This is because arriving early is a relatively innocuous outcome.

Relevance can explain loose talk

The previous examples show that a pragmatic listener embedding a “Combined” speaker model can explain intuitive patterns of inference. However, the most direct test is explaining speaker data: can our framework account for empirically-observed instances of loose talk?

An elegant experiment by van der Henst et al. (2002) provides particularly interesting data. Individuals on the street were asked the time, and responses were analyzed to determine what fraction of responses were rounded (a multiple of 5, such as 12:00 or 12:05). If speakers were following a “Truth-only” objective (under precise number semantics), exactly 20% of all responses should be rounded: speakers should only give a rounded time when it is literally true. However, van der Henst et al. find this is emphatically not the case: instead, the prevalence of rounding varies significantly depending on how the question is asked.

In particular, their Experiment 2 shows how manipulating the decision context systematically affects rounding behavior. Control participants were simply asked the time (“Do you have the time please?”), while Experimental participants were asked the time for the purpose of setting a watch (“My watch is going wrong. Do you have the time please?”). Participants asked the generic question almost always rounded (95.5%), while those asked the watch-setting
Table 3

Formalizing Relevance from van der Henst et al. (2002).

<table>
<thead>
<tr>
<th>Utterance $u$</th>
<th>:00</th>
<th>:01</th>
<th>:02</th>
<th>:03</th>
<th>:04</th>
<th>:05</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relevance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$U_R(u \mid w, A)$</td>
<td>Control</td>
<td>1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td></td>
<td>Watch</td>
<td>.5</td>
<td>.75</td>
<td>1</td>
<td>.75</td>
<td>.5</td>
</tr>
<tr>
<td><strong>Truth</strong> $U_T(u \mid w)$</td>
<td>-1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>-1</td>
<td>-1</td>
</tr>
<tr>
<td><strong>Cost</strong> $C(u)$</td>
<td>.5</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.5</td>
</tr>
</tbody>
</table>

Note. This table shows example utilities given a true time of 2 minutes past the hour (:02). We formalize the arguments advanced by van der Henst et al. by quantifying the decision-theoretic utility of utterances (Eq. 8), under the “Control” and “Watch” conditions respectively.

The authors suggest that when the inquirer’s decision problem is not explicit (as in the Control condition), speakers fall back to their general world knowledge to determine relevance. Since most activities (“appointments, TV programs, or university courses”) are scheduled to begin at a multiple of five, a person asking the time for unspecified purposes “will not be worse off” given a rounded time (van der Henst et al., 2002, p. 459). We formalize this intuition by defining a “Control” decision context, $A_{Control}$, with a positive payoff on the listener’s policy when told a time that lies within 5 minutes of the true time ($U_{Relevance} = 1$), and a negative payoff otherwise (-1). These utilities could be derived from a real-world decision problem that—for example—required the listener to arrive within a five minute window of a specified appointment time. In contrast, when setting one’s watch, knowing the time precisely is useful. In the $A_{Watch}$ context, we assume the decision-theoretic utility declines linearly as the answer becomes less accurate. Finally, van der Henst et al. (2002) suggest that rounded times require less processing effort; we formalize this by assigning them a lower cost $C(u)$ than unrounded times. An example payoff matrix is shown in Table 3, under the assumption that the true time is :02.

We then compare the distribution over utterances from two speaker models to that produced by human speakers. A purely truthful speaker always produces the precise time
Figure 4

*Decision-theoretic relevance can explain contextual acceptance of loose talk.*

![Graph showing the percentage of responses being multiples of 5 for different types of speakers under control and watch conditions.](image)

**Note.** Purely truthful speakers will always give the precise time, resulting in 20% of their responses being a multiple of five. A “Combined” speaker primarily optimizing for decision-theoretic relevance ($\lambda = .9$) will nearly always round the time under normal circumstances. However, when asked for the purpose of setting the listener’s watch (where a precise time has value), the percentage of rounding drops substantially. This pattern provides a close fit to empirical human data from van der Henst et al. (2002).

(yielding 20% of utterances which are multiples of 5), while a Combined speaker with $\lambda = .9, \beta_S = \beta_L = 10$ produces a human-like distribution: it chooses a rounded utterance 96% of the time under “Control” condition and 55% of the time under the “Watch” condition. Fig. 4 shows the results, and Fig. [B3] provides a parameter sensitivity analysis.

**Summary**

These results demonstrate that our framework can explain a number of seemingly contradictory results in the literature. Expanding our formal model of communication to incorporate the listener’s decision-making process allows this external context to determine appropriate standards of truthfulness. This, in turn, leads to context-dependent pragmatic strengthening or weakening of speakers’ literal meanings; and explains the variable acceptability of loose talk. In addition to explaining these outstanding theoretical puzzles, our framework makes a number of unique predictions about speaker’s behaviors when the two objectives diverge. In the next section, we introduce a novel extension to classic signaling games which allows us to
directly test these predictions.

**Signaling Bandits: decoupling truthfulness and relevance**

To determine how speakers weigh truthfulness and relevance, we need a controlled setting that decouples the epistemic and decision-theoretic utility associated with an utterance. Put simply, we need to create a diverse set of utterances: some that are true and relevant, some that are false and relevant, some that are true and irrelevant, and so on. We can then ask speakers to choose or endorse utterances with varying utilities to determine whether truthfulness or relevance (or some combination of the two) best explains their preferences.

In this section, we define a new signaling game (Skyrms, 2010) which satisfies this desideratum. We first review the structure of classic Lewis signaling games (Lewis, 1969), showing that they are not sufficient: epistemic and decision-theoretic utility are synonymous in these settings. We then describe contextual bandits, a decision setting studied in reinforcement learning (Lattimore & Szepesvari, 2020; Sutton & Barto, 2018). Finally, we combine Lewis games and contextual bandits to produce a new class of games, signaling bandits. Signaling bandits decouple the world state from the decision problem and fix the signal semantics to provide information about the world state. This allows us to independently vary the truthfulness (epistemic utility) and relevance (decision-theoretic utility) of utterances in order to test different theories.

**Lewis Signaling Games**

Lewis signaling games (Lewis, 1969) are two-player collaborative settings with a speaker and a listener (Fig. 5A). Following the notation introduced previously, such games are defined by a world state $w$, a set of actions available to the listener, $A \subseteq \mathcal{A}$, and a set of utterances available to the speaker, $\mathcal{U}$. There is one action $a^* \in A$ with a positive reward; other actions have zero reward. The world state $w$ implies the correct action $a^*$. The speaker knows $w$ but the listener does not. During gameplay, the speaker chooses an utterance $u \in \mathcal{U}$ and sends it to the listener. The listener updates their beliefs, $P_L(w \mid u)$, and uses the posterior to choose an action, $\pi_L(a \mid u, A)$. Lewis signaling games formalize the coordination problem underlying communication (Frank & Goodman, 2012; Krahmer & Van Deemter, 2012). However, the interplay of beliefs,
actions, and rewards is highly constrained. The state of the world \( w \) is synonymous with the correct action \( a^* \), and players are indifferent over other actions. This means that epistemic and decision-theoretic utility are perfectly correlated. As a result, it is not possible to discriminate the speaker objectives defined above (Eq. 5, 8, or 9): all predict the same distribution over utterances (Sumers, Hawkins, Ho, & Griffiths, 2021). For a richer decision-making setting, we turn to contextual bandits.

**Contextual Bandits**

Contextual bandits are an extension to the classic *multi-armed bandit* problem. Multi-armed bandits are single-player sequential games. In each round, the player takes an action and receives a scalar reward (Lattimore & Szepesvari, 2020; Sutton & Barto, 2018). Players seek to maximize their rewards, but are initially ignorant of the reward structure. Over multiple rounds, they must balance exploration (choosing a new action to learn its reward) with exploitation (choosing the most valuable known action). Contextual bandits extend this to a setting where actions are characterized by features, and rewards are defined with respect to these features.

For example, imagine learning to forage for mushrooms. Different varieties might be more or less tasty: green species might tend be delicious, while blue tend to be bitter. After eating a mushroom, associating payoffs with the color of the mushroom (a feature) rather than the mushroom itself (a specific action) allows knowledge to transfer to new settings (the next mushroom patch). Thus, unlike classic Lewis games, the payoff structure is correlated across contexts. The player must learn and exploit this correlation in order to choose optimal actions.

Formally, a function \( \phi \) associates features with each action: \( \phi : A \rightarrow \mathbb{R}^K \). Rewards are then defined as a function of these features: \( R : \phi(a) \rightarrow \mathbb{R} \). Thus, rather than learn about the reward of a specific action, players can learn about the reward of a *feature* which applies to many actions. Contextual bandits have been studied extensively in reinforcement learning (Abbasi-Yadkori et al., 2011; Chu et al., 2011; Li et al., 2010; Riquelme et al., 2018) and to a lesser degree for emergent communication (Donaldson et al., 2007). Here, we use them to study

\[11\] But see Qing and Franke (2015) for evidence in favor of action-oriented speakers in such settings.
Figure 5

Comparing Lewis signaling games to signaling bandits.

Note. (A) In classic Lewis signaling games (Frank & Goodman, 2012; Lewis, 1969), the world state is defined as the correct action (left). Speakers then use language to refer to this action (center). Knowledge of the world state is synonymous with knowledge of the correct action. Epistemic utility is thus perfectly correlated with decision-theoretic utility (right). (B) Signaling bandits defines an abstract world state in the form of feature-value pairings, which determine scalar payoffs associated with each action (left). Speakers use language to inform the listener about this world state (center). This breaks the correlation between epistemic and decision-theoretic utility: utterances may be false but useful, or true but not useful (right). Note that in this context, the utterance that maximizes decision-theoretic utility is false (“Spots are +2”).

Signaling Bandits

We combine the communication of Lewis games with the reward structure of contextual bandits to create a new class of games, signaling bandits (Fig. 5B). Unlike Lewis games, speakers no longer communicate concrete information (which action is correct). Instead, they communicate abstract information (how much features are worth). We now describe basic gameplay.

As in Lewis games, signaling bandits are two-player games with a speaker and a listener.
Each game is defined by a world state $w$, a set of all possible actions $\mathcal{A}$ and a set of speaker utterances $\mathcal{U}$. In each round, the listener faces a decision context $\mathcal{A} \subseteq \mathcal{A}$. However, unlike Lewis games, there is no single “correct” action. Instead, as in contextual bandits, each action has a scalar utility defined by a reward function which is independent of (and consistent across) contexts. While the reward function may be arbitrarily complex, in this work we use binary features and a linear reward function. Concretely, we assume features are indicator variables over actions:

$$\phi : \mathcal{A} \rightarrow \{0, 1\}^K \quad (10)$$

and rewards $R$ are linear over these features, parameterized by $w$:

$$R(a, w) = w^\top \phi(a). \quad (11)$$

This means the world state $w$ is a vector encoding the reward associated with each feature.

Fig. 6A depicts this visually: $w$ defines the value of individual colors and textures (table margins), which in turn determines the reward for each possible action in $\mathcal{A}$ (table contents). A listener with full world knowledge (every element of $w$) can calculate the exact rewards of every action and select the optimal action in any context. The value of an individual feature (one element of $w$) constitutes partial knowledge about the world.

The speaker helps the listener by providing such partial knowledge (Fig. 6B). In this work, we fix the semantics of signals: we take $\mathcal{U}$ to be a set of tuples of the form $(k, r)$ which specify a given feature and scalar value. As shown in Fig. 6C, these are messages like $(\text{Spots}, 1)$ or $(\text{Red}, 0)$. Speakers choose utterances and send them to the listener. The listener updates their beliefs according to Eq. 3, ruling out possible worlds where the feature takes on other values. The listener then chooses an action from the context $\mathcal{A}$ according to their posterior belief over rewards (Fig. 6C, Eqs. 6 and 7). This decouples the epistemic and decision-theoretic utility associated utterances, such that the two objectives yield different preferences over utterances (Fig. 6D).

Signaling bandits, like contextual bandits, are a class of games spanning a range of complexity. Lewis signaling games, which can be instantiated within signaling bandits (Sumers, Hawkins, Ho, & Griffiths, 2021), lie at the simplest end of this spectrum. More complex settings can incorporate imperative utterances and multiple decision contexts (Sumers et al., 2022).
**Figure 6**

*Signaling bandits formalities.*

---

**Note.** Signaling bandits combine Lewis signaling games with contextual multi-armed bandits. (A) The world $w$ is defined by correspondences between features and rewards (table margins), which combine additively to create possible actions $\mathcal{A}$ (table contents). In this world, green mushrooms tend to be tasty, while blue ones tend to be bitter. (B) One example context drawn from the world. A knowledgeable, cooperative speaker may produce an utterance before the listener takes an action. (C) Different utterances and their effects on the listener’s policy, $\pi_L(a | u, A)$ for $\beta_L = 1$. In this context, (Spots, +1) is true and relevant: it makes the listener likely to choose the spotted red mushroom. (Spots, +2) is *false* and relevant: exaggerating the value of spots increases the probability that the listener will choose it. Finally, (Green, +2) and (Red, 0) are true and *irrelevant*: they do not affect the listener’s policy. (D) Distribution over all 30 possible utterances induced by different speaker objectives for the decision context shown in B. Features (on the x-axis) are ordered in descending value; X’s on the diagonal mark true utterances. The purely truthful speaker ($\lambda = 0$, left) is insensitive to context and does not prioritize among true utterances, while the pure relevance speaker ($\lambda = 1$, right) readily exaggerates to obtain better actions. Combining the two ($\lambda = .5$, center) prioritizes utterances that are both truthful and relevant.
setting described here is a minimal extension to Lewis games, adding just enough complexity for meaningful differences to emerge between truthfulness and relevance. We describe these differences in more detail below, and return to future work using more complex signaling bandit settings in the General Discussion.

Using signaling bandits to study truthfulness and relevance

Why do signaling bandits—unlike Lewis’ signaling games—create a distinction between truthfulness and relevance? And how can this difference be used to test existing theories of communication? We now examine these questions in detail, highlighting how game dynamics create divergent predictions.

The basic innovation is that a single world state (the margins of the table in Fig. 6A) can be used to construct many decision contexts (any set of the actions from $A$, the contents of the table in Fig. 6A). Utterance semantics are then fixed, and describe this abstract world state (they provide information about $w$, i.e., “Green is worth +2”) rather than simply referring to a concrete action (c.f. “Green”, Frank & Goodman, 2012). Thus, utterances are true or false regardless of the context: their epistemic utility depends only on the world state $w$ (Eq. 5). However, their relevance is context-dependent: their decision-theoretic utility depends on the specific decision problem $A$ (Eq. 8). This decouples truthfulness and relevance (Fig. 5).

To see this, consider the context and utterances shown in Fig. 6C. The utterance (Spots, 1) is both true and relevant: the listener’s decision policy skews towards the spotted red mushroom, improving it over a random choice. However, utterances can also be false and relevant: for example, the utterance (Spots, 2) exaggerates the value of spots. This utterance induces false beliefs about the underlying world $w$ (and thus has negative epistemic utility), but yields a high probability of the listener choosing the optimal action (and thus has positive decision-theoretic utility). Finally, utterances may be true and irrelevant. It is straightforward to see that (Green, 2) is not relevant, because the green feature is not present in this context. A more subtle example is the utterance (Red, 0). Red is the most common feature in this context, and so its value could plausibly be considered important. However, under the assumption of a uniform prior over feature values, learning that a feature is worth 0 does not change its expected utility. As a result, it too is
irrelevant: hearing this utterance does not affect the listener’s policy. In this decision context, both (Green, 2) and (Red, 0) reduce the listener’s uncertainty about the world \( w \) (positive epistemic utility), but do not affect the listener’s actions (zero decision-theoretic utility).

How does this decoupling help us test competing theories? Fig. 6D illustrates how speaker objectives now induce different preferences over utterances. Each grid represents the 30 possible utterances (6 features \( \times \) 5 values), with X’s indicating true utterances: (Green, 2) is a true utterance, while (Green, 1) and (Spots, 2) are not. Shading indicates the distribution over utterances resulting from each objective in the context shown in Fig. 6B. A purely truthful speaker weighs only the epistemic utility of utterances. Thus, it eschews false utterances but is indifferent over true ones, placing a uniform probability over them. At the other extreme, a pure relevance speaker chooses utterances proportional to their decision-theoretic utility, with no regard for truthfulness. Thus it favors exaggerating the “Spots” or “Stripes” features, or ascribing positive utility to the neutral “Red” feature. Combining the two objectives yields an entirely new prediction: a strong emphasis on the (Spots, 1) utterance, which compromises between truthfulness and decision-theoretic relevance.

These distinctions are not possible in traditional Lewis signaling games, which align the two objectives (Sumers, Hawkins, Ho, & Griffiths, 2021).

In the following sections, we use behavioral experiments to test these theoretical predictions. Our first experiment gives participants access to the full utterance space. We hypothesized that both truthfulness and relevance are important constraints on communication, such that our Combined model would predict response patterns better than either alone. Our second experiment focuses on the phenomenon of “loose talk,” asking participants to endorse false but relevant utterances. This affords a strong test of the primacy of these different objectives:

---

12. Note that these particular decision-theoretic utilities are a function of the context. To see this, we can consider a context consisting of three spotted mushrooms: one red, one green, and one blue. In this context, both (Spots, 1) and (Spots, 2) are irrelevant (because they do not disambiguate between actions), while the true utterance (Green, 2) is now relevance-maximizing.

13. Similar tension between truthfulness and relevance is evident even in real-world information about mushrooms. Textbooks on mycology bias towards truthfulness, providing a comprehensive but academic perspective on fungi (Deacon, 1997). In contrast, practical foraging guides bias towards relevance, dramatically simplifying the underlying biology in favor of useful heuristics to avoid getting sick (Hyman, 2021).
Relevance Theory (Sperber & Wilson, 1986; Wilson & Sperber, 2002b) suggests that participants should always endorse these; truth-based formalisms (Goodman & Frank, 2016; Lewis, 1969) suggest they should never; and our Combined model predicts that endorsement will vary as a function of decision-theoretic utility.

**Experiment 1: Is truthfulness or relevance alone sufficient?**

Our first experiment asks whether truthfulness or relevance alone can explain participants’ utterance choices. We give speakers access to the full semantic space in the signaling bandits game introduced in the previous section, allowing them to choose from 30 possible utterances. While previous approaches have emphasized either truthfulness or relevance, we hypothesized that participants would be sensitive to both, such that their pattern of responses would be best explained by our combined objective (Eq. 9).

To put the participant in the role of a cooperative and knowledgeable speaker, we used a cover story situating them as a tour guide giving visitors information about local mushroom species. Participants were first trained and tested on the world state (Fig. 6A) which was randomized across participants (Fig. C1). After passing a comprehension check, participants then provided advice to tourists in different decision contexts (Fig. 6B), instantiated as visiting different mushroom patches (Fig. C2).

We used two control conditions and one experimental condition in a fully between-subjects design. Our control conditions manipulated participants’ objectives to test their understanding of epistemic and decision-theoretic utility respectively. Participants in the “Truth-biased” condition were instructed to provide true facts only, testing whether participants understood the truthfulness of utterances. Participants in the “Relevance-biased” condition were instructed to ensure tourists picked good mushrooms (and that lying was allowed), thus testing whether the participants’ model of the listener’s decision-making matched ours (Eq. 7). Finally, the experimental “Unbiased” condition was used to test participants’ default behavior. Participants

---

14 This study was approved by the Princeton IRB (protocol #10859). Experiment design and analyses were pre-registered at https://aspredicted.org/SPK_Z7Y All materials, code, and data are available at https://github.com/tsumers/relevance The experiment can be seen at https://signaling-bandits.herokuapp.com
Figure 7

Results from Experiment 1.

Note. Top: participant responses across conditions for the example context. X’s indicate true utterances. Unbiased participants were sensitive to both truthfulness and relevance: they strongly favored the “Spots, 1” utterance, which is both true and relevant. In contrast, Truth-biased participants additionally favored the “Red, 0” utterance (true but not relevant), while Relevance-biased participants chose the “Spots, 2” utterance (false but highly relevant). Bottom: to visualize the distribution of the \( \lambda \) parameter within each condition, we fit the Combined model to individual participants. Histograms show participant-level \( \lambda \) MLEs, while dashed lines and bold text show condition-level MLEs. Participant-level MLEs confirm that virtually all Unbiased participants considered both utilities; instructing participants to follow either biased them towards that theoretical model.

in this condition were given no instructions about their objective. Our key hypothesis was that these Unbiased participants would optimize for both truthfulness (epistemic utility) and relevance (decision-theoretic utility), and thus be best explained by the “Combined” model.
Methods

Participants

We recruited 301 participants using Prolific (www.prolific.co). Participants were required to be fluent in English, possess an approval rating of 95%, and be located within the United States, United Kingdom, or Ireland. They were paid $2, with an additional completion bonus of $2 if they passed a pre-experiment comprehension check. 16 participants failed this comprehension check and 12 failed attention checks during the experiment, leaving a final sample size of 273. On average, participants spent about 15 minutes on the experiment (M=15.71, SD=6.02) and earned an hourly wage of $14.87.

Stimuli

The experiment was designed as a direct implementation of signaling bandits as described above. We thus used the reward structure shown in Fig. 6A (see Fig. C1 for screenshots).

Our experimental trials used decision contexts consisting of 3 distinct actions, $A \in [A]^3$, for a total of 84 different contexts. We divided these into 3 sets of 28 and randomly assigned participants to one set. On each of these trials, participants were shown a tourist visiting a particular mushroom patch (i.e., a decision context $A$). They were asked to choose an utterance of the form “<feature> is <value>,” using drop-down menus to select one of six features (Green, Red, Blue, Spotted, Solid, Striped) and one of five values (-2, -1, 0, 1, 2). The feature-value mapping, trial ordering, and menu ordering were randomized (see Appendix C for details).

Procedure

Participants began with an instruction phase, where they were given the cover story about providing advice to tourists. They were first taught the true world state $w$ (Fig. 6A). They were then taught signaling bandits dynamics. Specifically, they were told that tourists always visited a single mushroom patch (consisting of three mushrooms) and chose a single mushroom from it. Tourists knew nothing about the mushrooms, but the participant could produce a single utterance (chosen from the 30 possible feature-value tuples) before the tourist picked one. The tourists’ decision-making process (Eq. 7) was not specified.
Participants in the Truth- and Relevance-biased control conditions were shown an additional page specifying their responsibilities as a tour guide. Truth-biased participants were told their job was “teach tourists facts about mushroom features,” and instructed that “it does not matter what mushrooms they choose.” In contrast, Relevance-biased participants were told their job was to “ensure tourists choose tasty mushrooms,” and that “it does not matter if you tell the truth or not” (see Fig. C3 for screenshots). Finally, in the experimental Unbiased condition, participants were given no instructions about their objective.

After reading the instructions, participants took a comprehension quiz consisting of sixteen multiple choice questions. To ensure participants had learned the world state \( w \), they were required to correctly provide the value of all six individual feature values and three specific actions. To ensure they understood the game dynamics, they were asked seven questions about the experiment itself. Two questions were condition-dependent: Truth- and Relevance-biased participants were asked about their objective, while Unbiased participants were asked neutral questions about the game dynamics. Participants were required to answer all 16 questions correctly, but were given three opportunities to do so and could review the instructions between each attempt. If they failed three times, the experiment ended early.

Participants who passed began the experiment itself, consisting of 28 trials and 8 attention checks (for details, see Appendix C). To avoid memory confounds, participants could view the full world state at any time (Fig. C1).

Results

Analysis of participant behavior shows that our objective manipulation strongly affected participants’ behaviors. Across all contexts, rates of truth-telling varied significantly by condition \((F(2, 270) = 49.40, p < .0001)\). Truth-biased participants chose truthful utterances 95% of the time; Unbiased 85%, and Relevance-biased just 67% of the time.

Control conditions biased participants towards their respective theoretical models, confirming that participants understood and could independently optimize epistemic and decision-theoretic utility. Crucially—and as predicted—Unbiased participants jointly optimized truthfulness and relevance, rather than either alone. Fig. 7 shows the empirical distribution of
Table 4

*Maximum likelihood estimates in Experiments 1 and 2.*

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Condition</th>
<th>N</th>
<th>$\lambda$</th>
<th>$\beta_S$</th>
<th>$\beta_L$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Free Choice</td>
<td>Truth-biased</td>
<td>87</td>
<td>.35</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Unbiased</td>
<td>95</td>
<td>.55</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Relevance-biased</td>
<td>91</td>
<td>.85</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>2: Forced Choice</td>
<td>Truth-biased</td>
<td>71</td>
<td>.15</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Unbiased</td>
<td>78</td>
<td>.75</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Relevance-biased</td>
<td>79</td>
<td>.90</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

utterances chosen by participants for the example context. Qualitative comparison with Fig. 6D highlights several important trends. First, control conditions affected participants’ choice of utterances in this context. For example, Relevance-biased participants readily used false utterances to achieve a desired action: they sent \(\text{Spots, }+2\) as well as \(\text{Red, }+1\) and \(\text{Red, }+2\). This demonstrates that participants understood and could optimize for decision-theoretic utility independent of epistemic utility. Second, Unbiased participants clearly followed a combination of the two objectives: they overwhelmingly favored an utterance that was truthful and relevant, but not necessarily relevance-maximizing: \(\text{Spots, }+1\).

**Parameter estimates across conditions**

We now analyze the $\lambda$ parameter, which determines the relative weight of the truthfulness and relevance objectives in Eq. 9. Higher values ($\lambda \approx 1$) indicate that participants’ response patterns were best explained by relevance (decision-theoretic utility), while lower values ($\lambda \approx 0$) indicate their responses were best explained by truthfulness (epistemic utility).

We implement our model in WebPPL (Goodman & Stuhlmüller, 2014) and use a grid search to infer model parameters. Our full grid consisted of $\lambda \in [0, 1]$ in steps of .05 and $\beta_S, \beta_L \in [1, 10]$ in steps of 1. Maximum likelihood estimates across conditions (Table 4) show the effect of our manipulation: $\lambda_{\text{Truth-biased}} = .35$, $\lambda_{\text{Unbiased}} = .55$, $\lambda_{\text{Relevance-biased}} = .85$. A model
Table 5

Model comparison for the Unbiased condition.

<table>
<thead>
<tr>
<th>Model</th>
<th>Log Bayes Factor</th>
<th>Experiment 1</th>
<th>Experiment 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Truthfulness Only</td>
<td>1475</td>
<td>846</td>
<td></td>
</tr>
<tr>
<td>Relevance Only</td>
<td>1508</td>
<td>276</td>
<td></td>
</tr>
</tbody>
</table>

comparison across conditions confirmed these differences are significant (see Appendix C). In addition to these condition-level comparisons, we separately fit participant-level MLEs to visualize the distribution of the $\lambda$ parameter within conditions (Fig. 7, bottom).

Successful manipulation of the $\lambda$ parameter confirms that participants understood epistemic and decision-theoretic objectives, and could optimize them independently. Having established this, we focus on the Unbiased condition and ask whether the two component models (truthfulness or relevance alone) are sufficient to explain participant responses.

Model comparisons for Unbiased participants

Our primary hypothesis is that response patterns in the Unbiased condition will be best explained by a combination of epistemic truthfulness and decision-theoretic relevance. We perform a model comparison using the grid search described above, restricting our results to $\lambda = 1$ to obtain the “Relevance only” model and $\lambda = 0$ to obtain the “Truthfulness only” model. We use Bayes factors to compare marginal likelihoods for these models. Evidence for the “Combined” model is extremely strong, with Bayes factors in favor of the “Combined” model $> 1 \times 10^{1400}$; we report log Bayes factors in Table 5. Likelihood ratio tests comparing the “Combined” model to the two nested models yield similar results (vs. “Relevance only”: $\chi^2(1) = 2956.62, p < .0001$; vs. “Truthfulness only”: $\chi^2(2) = 3026.66, p < .0001$). These results confirm our hypothesis: evidence overwhelmingly favors the Combined model to explain Unbiased participants’ utterance choices. These participants followed a combination of epistemic truthfulness and decision-theoretic relevance, rather than either objective independently.
Figure 8

Model performance for Unbiased condition in Experiment 1.

![Graph showing model predictions for Unbiased condition](image)

Note. Variance explained for Unbiased condition. We compare component models (“Truthfulness” and “Relevance” only) to the full model (“Combined”). Predictions are made with MLE parameters in Table C1. Independently, truthfulness and relevance are insufficient constraints to explain humans’ preferred choice of utterance. Fitting an additional cost term improves $r^2$ to .81 (Fig. C6).

We next use the MLE parameters (as well as parameters estimated for the two component models, Table C1) to generate predictions for each model and compare them against human behavior (Fig. 8). It is clear that both truthfulness and relevance are important constraints: neither component alone is sufficient to explain response patterns. Inspection of the residuals (Fig. C5) shows that the model overpredicts usage of negative utterances such as (Blue, -2), and underpredicts usage of positive ones such as (Green, +2). This suggests that negative utterances may be more costly to produce or comprehend: intuitively, they tell the listener what to avoid, rather than what to choose. This suggests participants balanced “processing effort” against decision-theoretic utility (Jara-Ettinger & Rubio-Fernandez, 2021b; Sperber & Wilson, 1986; Wilson & Sperber, 2002a). Following standard practice in RSA models (Goodman & Frank, 2016), we added a cost term based on the utterance valence, which improved the model fit (see Appendix C, Figs. C6 and C7).

Finally, we hypothesized that maximizing the listener’s decision-theoretic utility would require cognitive effort. To test this, we predicted individual participants’ per-trial response times (M=12.29 seconds, SD=11.14) from their estimated $\lambda$ parameter. We used a mixed-effects linear
regression, with a fixed effect for the participant’s inferred $\lambda$ and random effects for each participant (Barr et al., 2013). The effect of $\lambda$ was positive and significant ($\beta = 2.99, t(271) = 3.32, p < .01$; see Table C2). This suggests that maximizing the listener’s utility took more effort than simply choosing a truthful utterance.

Discussion

Our results provide strong evidence that participants’ decisions were guided by a combination of truthfulness and decision-theoretic relevance. Our control conditions (Truth- and Relevance-biased) demonstrate that participants understood these distinct objectives and were capable of optimizing them independently: for example, Relevance-biased participants condition readily sent false messages exaggerating features’ values. However, few participants in the Unbiased condition did so: instead, they preferred true utterances that maximized utility.

Additionally, we find evidence that reasoning about the listener’s decision problem may be cognitively expensive. Maximizing decision-theoretic utility correlated with longer response times (Table C2), and residual analysis suggests that participants favored positive-valued utterances which may require less “processing effort” (Figs. C5-C7). We return to more nuanced formulations of decision-theoretic processing costs, such as the entropy of the listener’s decision policy, in the General Discussion.

Experiment 1 establishes that truthfulness and relevance may be effectively combined to explain participants’ production patterns. However, this leaves an important question unanswered: how, exactly, do the two objectives relate to each other? Does one take priority over the other? Experiment 2 was designed to address this question by putting the two objectives in conflict.

**Experiment 2: Are false utterances ever acceptable?**

Experiment 1 demonstrated that truthfulness and relevance operate as independent constraints on participants’ utterance choices. Combining these objectives with a simple convex model produces strong quantitative fits to participant behaviors. However, while these results validate the importance of both truthfulness and relevance, they cannot determine which is
primary. Does one maxim outweigh the other?

Asking participants to endorse false but useful utterances (which lie in the lower-right quadrant of Fig. 1) provides a particularly strong litmus test of different formal models. Pure truth-oriented models (Goodman & Frank, 2016; Stalnaker, 1978) predict that participants should never endorse them. In contrast, Relevance Theory suggests that relevance is primary and truthfulness is an epiphenomenon (Sperber & Wilson, 1986; Wilson & Sperber, 2002b). Under this perspective, participants should endorse relevant utterances irrespective of their truthfulness. Contrary to both of these theories, our “Combined” model predicts that participants will exhibit a graded tradeoff between the two objectives. Concretely, we hypothesized that participants would refuse to endorse false utterances that provide little decision-theoretic utility, but that endorsement rates should rise monotonically as decision-theoretic utility increases.

We designed Experiment 2 to test this hypothesis. We asked participants to endorse true and false utterances while systematically varying their decision-theoretic utility. We predicted that participants would endorse a false utterance when—and only when—that lie carried sufficient decision-theoretic utility (and conversely, decline to endorse a true utterance if it had sufficiently negative decision-theoretic utility). As with Experiment 1, such a response pattern would be best explained by our “Combined” model.15

Methods

Participants

Experiment 2 used the same qualification and compensation structure as Experiment 1. We recruited 300 participants; 25 failed the comprehension quiz and 47 failed attention checks, leaving a final sample size of 228. On average, participants spent about 16 minutes on the experiment (M=16.47, SD=6.84) and earned an hourly wage of $13.96.

15 This study was approved by the Princeton IRB (protocol #10859). Experiment design and analyses were pre-registered at https://aspredicted.org/9MD_THB
Figure 9

*Schematic of Experiment 2 trials.*

**A**

Allen is visiting this mushroom patch. Would you say: 

"Spots are +1"

<table>
<thead>
<tr>
<th>Say</th>
<th>Say</th>
</tr>
</thead>
<tbody>
<tr>
<td>nothing</td>
<td>&quot;Spots are +1&quot;</td>
</tr>
</tbody>
</table>

Epistemic ✔
Decision-theoretic ✔

**B**

Anne is visiting this mushroom patch. Would you say: 

"Red is +1"

<table>
<thead>
<tr>
<th>Say</th>
<th>Say</th>
</tr>
</thead>
<tbody>
<tr>
<td>nothing</td>
<td>&quot;Red is +1&quot;</td>
</tr>
</tbody>
</table>

Epistemic ✘
Decision-theoretic ✔

**C**

James is visiting this mushroom patch. Would you say: 

"Spots are +1"

<table>
<thead>
<tr>
<th>Say</th>
<th>Say</th>
</tr>
</thead>
<tbody>
<tr>
<td>nothing</td>
<td>&quot;Spots are +1&quot;</td>
</tr>
</tbody>
</table>

Epistemic ✔
Decision-theoretic ✘

*Note.* Experiment 2 trial structure. In each trial, participants were given a specific utterance-context pair and used a slider to indicate their preference between saying it or staying silent. (A) An “aligned” utterance that is true and has positive decision-theoretic utility. (B) A “conflicted” utterance that is false but has positive decision-theoretic utility: “Red” is worth 0, but hearing this utterance makes the listener less likely to choose the negative blue mushroom. (C) A “conflicted” utterance that is true but has negative decision-theoretic utility. The “Spots” feature is actually worth +1, but here co-occurs with the highly negative “Blue” feature. Hearing “Spots are +1” induces true beliefs that lead the listener to take a suboptimal action, reducing their expected utility below random chance. See Fig. [D1](#) for screenshots of the experiment.

**Stimuli**

The basic stimuli were the same as Experiment 1. However, the trial structure was changed from selecting an utterance to endorsement of a specific utterance. On each trial, participants were given an utterance (for example, “Spots are +1”) and chose between saying it or saying nothing. Participants used a slider to indicate their response: the left side was labeled “Definitely stay silent” and the right “Definitely say [utterance].” (Fig. [9](#)). Slider responses were recorded as integer values ranging from [0, 100].

Each trial consisted of a context-utterance pair. We first identified 72 pairs that put the two objectives in *conflict*. These consisted of false utterances (negative epistemic utility) that—in the paired context—were relevant (positive decision-theoretic utility); or true utterances (positive
epistemic utility) which were not (negative decision-theoretic utility). We balanced this with a sample of 72 aligned pairs (i.e., negative epistemic and decision-theoretic utility; or positive epistemic and decision-theoretic utility). Fig. 9 shows example aligned and conflicted trials. This gave a total of 144 possible trials. We grouped these trials by feature and truthfulness, then sorted each group by decision-theoretic utility. We then round-robin assigned them into four sets of 36. This ensured that each participant saw a comparable distribution of utterances, features, and corresponding utilities, mitigating the risk of individual participants seeing sustained spurious correlations (such as the negative “blue” feature consistently correlating with the positive “spotted” feature, Fig. 9C). We then randomly assigned participants to a set, and randomized the trial ordering for each participant. As in Experiment 1, we included 8 attention checks which were held constant across all participants. Appendix D contains additional details.

**Procedure**

Our procedure was identical to Experiment 1, except that the interface was changed to the slider endorsement described above and participants gave judgments for 36 experimental trials instead of 28.

**Results**

As anticipated, the endorsement paradigm forced a sharper separation of speaker objectives. Our control conditions illustrate this: Truth- and Relevance-biased participants adhered closely to the respective theoretical models. Truth-biased participants always endorsed true statements and refused to endorse false ones, while Relevance-biased participants endorsed statements proportional to their decision-theoretic utility while ignoring their truthfulness (Fig. 10, top left and right). This confirms that participants were unambiguously aware of the epistemic utility (truth or falsehood) of utterances and sensitive to even small fluctuations in their decision-theoretic utility across contexts.

As predicted, Unbiased participants did not simply prioritize one objective over the other. Instead, they demonstrated sensitivity to both: endorsement primarily followed decision-theoretic utility, but clearly distinguished between true and false utterances. Their preference for true
**Figure 10**

*Results from Experiment 2.*

![Graph showing results from Experiment 2.](image)

*Note.* Experiment 2 asked participants to endorse utterances with varying truthfulness and relevance. Top: response patterns across conditions. Scatterplots show individual responses, solid lines are locally-weighted regressions, and dashed lines are MLE predictions. Truth- and Relevance-biased participants endorsed utterances proportional to their epistemic or decision-theoretic utility alone. This confirmed that they understood the truthfulness of utterances and their effects on listener behaviors. As predicted, Unbiased participants followed a mix of epistemic and decision-theoretic objectives. Participants endorsed true utterances proportional to their decision-theoretic utility. They were less willing to endorse false utterances, indicating a sensitivity to epistemic utility—but endorsement nonetheless increased steadily with decision-theoretic utility. Thus, neither objective strictly dominates: when put in conflict, participants made a graded tradeoff between the two. Our MLE captures these trends, although it overpredicts endorsement of true-but-misleading utterances. This suggests that epistemic utility may in fact be asymmetric (see the General Discussion). Bottom: MLEs for the $\lambda$ parameter across conditions. Histograms show the distribution of participant-level MLEs within each condition (estimated by fitting the Combined model to each participant independently), while dashed lines and bold text show the condition-level MLEs (estimated by fitting the Combined model to the condition as a whole). The endorsement paradigm made it difficult to satisfy both objectives.
utterances is visible as the horizontal distance between true and false endorsement curves in the top center panel of Fig. [10] Participants were willing to forego decision-theoretic utility to avoid telling a lie, indicating that they valued truthfulness independent of relevance.

To assess quantitative results, we first adapt our model to the endorsement paradigm, then follow the same model comparison procedure as Experiment 1. We formalize the choice of not saying anything by modeling a “silent” utterance with no effect on the listener’s beliefs. We modify the “truthfulness” objective (Eq. 5) and set the epistemic utility of the “silent” choice to zero. The “relevance” objective naturally (Eq. 8) reduces to the expected utility of a random choice from the context. We then model each trial as a binary choice between two utterances: the presented utterance and this “silent” utterance. To account for noise in the participants’ slider-based judgments, we add Gaussian noise $\epsilon \sim N(0, 30)$ around the model prediction before evaluating model likelihood with respect to the slider judgment. We set the variance of this noise distribution based on pilot data.

**Parameter estimates across conditions**

We again use a gridsearch over model parameters and analyze the the $\lambda$ parameter to assess which objectives participants followed. We used the same grid as Experiment 1: $\lambda \in [0, 1]$ in steps of .05 and $\beta_S, \beta_L \in [1, 10]$ in steps of 1. Maximum likelihood estimates (Table 4) for control conditions confirm that Experiment 2’s forced-choice paradigm provides a stronger separation of the two objectives: $\lambda_{Relevance-biased} = .90, \lambda_{Truth-biased} = .15$. Our forced-choice paradigm particularly affected the Belief-biased condition, as participants no longer had the flexibility to choose utterances that were truthful and also optimized relevance. Remarkably, participants in the Unbiased condition still balanced the two: $\lambda_{Unbiased} = .75$. As with Experiment 1, manipulation checks confirmed our manipulation had significant effects on participants’ response patterns (Appendix D). Again, following Experiment 1, we then fit the Combined model to individual participants in order to visualize the distribution of $\lambda$ across participants within each condition. Unbiased participants’ MLEs followed a roughly bimodal distribution, with more concentration at the extremes (Fig. [10] bottom center)—notably different from Experiment 1, which followed a more uniform distribution (Fig. [7] bottom center). While
many individual participants were best explained by the “Combined” model, a nontrivial fraction did adhere strongly to one objective.

**Model comparisons for Unbiased participants**

We followed the same procedure as Experiment 1 to perform a model comparison. Our results again provided extremely strong evidence for the “Combined” model, with Bayes factors exceeding $1 \times 10^{200}$ (Table 5). Likelihood ratio tests comparing the “Combined” model to the two nested models yield similar results (vs. “Relevance only”: $\chi^2(1) = 556.36, p < .0001$; vs. “Truthfulness only”: $\chi^2(2) = 1701.03, p < .0001$).

There are two possible reasons for the “Combined” model’s success. The first is that individual participants followed one of the component models (truthfulness- or relevance-only), but differed in which utility they prioritized. The “Combined” model would then be the best explanation for the condition-level results, but any individual participant would be better explained by a simpler truth- or relevance-only model. In contrast, it might be that individual participants followed the “Combined” model and made a graded tradeoff between the two utilities. To test these two hypotheses, we conducted an additional model comparison allowing the $\lambda$ parameter to vary by participant. We compared a simpler “Bimodal” model assuming $\lambda \sim \{0, 1\}$ against a more complex “Combined” model allowing $\lambda \sim \text{Uniform}(0, 1)$. The “Bimodal” model instantiates the hypothesis that individuals followed one of the component models, whereas the “Combined” model suggests that individuals weighed both utilities. We used annealed importance sampling to compare the two (Grosse et al., 2016) and found extremely strong evidence that the “Combined” model provided a better fit to the data (Bayes factor $6.27 \times 10^{13}$; see Appendix D for details). This confirms that individual participants made a graded tradeoff between truthfulness and relevance: the “Combined” model reflects the actual utility function followed by individuals, not simply an overall trend from mixing heterogeneous populations.

---

16 We thank an anonymous reviewer for suggesting this analysis.
Discussion

The forced-choice paradigm allowed us to pit truthfulness directly against relevance. Our control conditions reflect this: biasing participants to focus on either truthfulness or relevance moved them strongly towards that objective. Most importantly, Unbiased participants followed the graded tradeoff predicted by our Combined model: participants showed a willingness to endorse false utterances, but a preference for true ones. This preference—visible as a horizontal offset between endorsement of true and false statements in Fig. 10—captures the “price” that participants put on falsehoods, measured in decision-theoretic utility. This graded tradeoff combines a sensitivity to decision-theoretic utility with a preference for truthfulness (Abeler et al., 2019)—highlighting the need for integrated models considering both utilities.

General Discussion

In this work, we introduced a new speaker model synthesizing belief- and action-based theories of communication. Expanding the scope of belief-oriented models to encompass real-world action allows a speaker to value both the epistemic and decision-theoretic consequences of speech acts. This framework allowed us to reconcile formal definitions of truthfulness and relevance, capturing a range of examples discussed in the literature. To test our theory, we introduced a new signaling game (Lewis, 1969; Skyrms, 2010) and conducted a pair of experiments. Experiment 1 confirmed that speakers follow both truthfulness and relevance, while Experiment 2 showed that participants make a quantitative, graded tradeoff between the two. This graded tradeoff challenges central tenets of both belief- and action-oriented theories of communication, which hold that one utility is primary. Instead, we find that epistemic and decision-theoretic utilities must be considered in tandem (Franke, 2009).

Implications for psychology

What, then, are the implications of our work? First, our results suggest that speaker behavior results from an interaction between different objectives rather than a single primary goal. Truthfulness can be seen as a preference (Abeler et al., 2019) rather than the foundation of human communication. Formal models emphasizing truthfulness (Lewis, 1969; Stalnaker, 1978)
may be best understood as normative or deontological accounts, with real-world speakers deviating for a variety of reasons—such as rounding times to reduce processing costs (Gibbs Jr & Bryant, 2008; van der Henst et al., 2002), or spreading fake news as a result of failing to fact-check (Pennycook et al., 2021; Pennycook & Rand, 2021).

Similarly, our approach challenges several claims from Relevance Theory. Sperber and Wilson (1986) describe relevance in purely cognitive terms, but our theoretical framework derives it from external decision-theoretic utility (P. Parikh, 1992; R. Parikh, 1994; van Rooij, 2003). Further, our empirical findings show that participants independently value the truthfulness of an utterance, undermining the idea that communication is best understood in terms of relevance alone (Wilson & Sperber, 2002b). Despite these discrepancies, we suggest that our framework may be understood as an extension of the ideas proposed by relevance theorists. Expanding the scope of our formal theory beyond discourse allowed us to quantify “positive cognitive effects” in terms of real-world behaviors. Viewed in this light, our results support the basic claim that relevance is a crucial consideration which may outweigh truthfulness under some circumstances.

Finally, deriving relevance from decision theory may explain the origin of the “question under discussion” (QUD, Benz & Jasinskaja, 2017; van Kuppevelt, 1995; Roberts, 2012). As noted by Roberts (2012), “discourse goals” are often determined by external “domain goals.” Decision theory formalizes this process: question-asking agents can use the “value of information” (Savage, 1954) to prioritize inquiry about decision-relevant information (P. Parikh, 1992; R. Parikh, 1994; van Rooij, 2003). Conversely, speakers can use the same principles to select information that maximally improves the listener’s decision-making (Benz, 2006; Benz & Van Rooij, 2007). This can be seen when speakers “go beyond” the literal content of a question and supply additional decision-relevant knowledge (Hawkins et al., 2015). For example, asked if they accepted credit cards, restaurant owners answered the question and then spontaneously informed the caller about unexpected closures: “Uh, yes, we accept credit cards. But tonight we are closed” (Clark, 1979, p. 466). The restaurant’s open hours are irrelevant to the caller’s literal question, yet clearly relevant to the questioner’s implicit plan to come for a meal. Decision-theoretic relevance thus formalizes the real-world goals atop the QUD hierarchy (Benz & Jasinskaja, 2017; Roberts, 2012).
Implications for artificial intelligence

Beyond psychology, our work offers a new perspective for AI systems. Autonomous agents that understand language (Luketina et al., 2019; Tellex et al., 2020) and reason pragmatically (Fisac et al., 2020; Golland et al., 2010; Hadfield-Menell et al., 2016; Jeon et al., 2020; Milli et al., 2017; Sumers et al., 2022; Wang et al., 2016) are an active research area, but these works typically assume that humans act to maximize decision-theoretic utility alone. Our findings suggest that humans prefer to produce utterances that are relevant and true, suggesting that pragmatic speaker models should reflect these biases (Shah et al., 2019).

A second line of AI research develops large language models (LLMs; T. Brown et al., 2020; Chowdhery et al., 2022; Thoppilan et al., 2022). LLMs are trained on massive corpora scraped from the internet, leading them to produce fluent but often harmful language (Bender et al., 2021; Weidinger et al., 2021). This has led to calls to integrate Gricean maxims into such systems to promote normative cooperative discourse (Kasirzadeh & Gabriel, 2022). Computational models of human communication, such as the one presented here, can help bridge the gap. For example, future work could integrate decision-theoretic relevance into LLMs by asking them to reason through the real-world behavioral consequences of an utterance (Ahn et al., 2022; Huang et al., 2022) and evaluate the resulting utilities (Jin et al., 2022) prior to producing it.

How do speakers model listener decision making?

Measures of decision theoretic utility (Eq. 8) typically assume speakers know both the listener’s decision problem (\(A\), the set of available actions) and reason perfectly about their decision-making process (Benz, 2006; Benz & Van Rooij, 2007; Gmytrasiewicz & Durfee, 2001; P. Parikh, 1992; R. Parikh, 1994; Qing & Franke, 2015, *inter alia*). Are these assumptions reasonable? We first consider justifications for making them, then discuss how they may break down in real-world communication.

First, it has long been recognized that language comprehension relies on “common sense” knowledge about real-world scenarios (Sanford & Garrod, 1981; Schank & Abelson, 1977). This is equivalent to our assumption that the speaker knows the listener’s decision problem: for example, Grice’s classic “out of petrol” example of relevance (Grice, 1975) presumes that the speaker and
listener know that cars need petrol to drive, and garages typically sell petrol.\footnote{In artificial agents, such information can be supplied as an external knowledge base (Gmytrasiewicz & Durfee, 2001; Tambe, 1997) which is used to derive situation-specific payoff matrices.} Second, both adults (Baker et al., 2017; Baker et al., 2009; Jara-Ettinger et al., 2016) and children (Gergely et al., 2002; Gergely et al., 1995) use expectations of rational action together with situational constraints to drive sophisticated social inference. This theory of mind supports both learning (Aboody et al., 2022) and teaching (Bass et al., 2019; Ho, Saxe, et al., 2022), allowing even young children to select task-relevant communicative acts (Gweon & Schulz, 2019) which maximize the listener’s decision-theoretic utility (Bridgers et al., 2020; Gweon, 2021). Decision-theoretic relevance applies the same theory of mind to everyday discourse, assuming the speaker models the listener as a rational agent. Implicit world knowledge supplies the set of possible actions and associated utilities, while theory of mind supplies the listener’s decision process (Davis & Jara-Ettinger, 2022).

The speaker’s knowledge need not be perfect. Just as RSA has been extended to account for lexical uncertainty (Degen et al., 2020; Smith et al., 2013; Wang et al., 2016), our formulation can incorporate uncertainty by marginalizing over decision problems, listener belief states, or true world states. This allows a speaker to reason about—for example—decision problems that might arise in the future (Sumers et al., 2022). And in everyday discourse, a speaker may have partial (Vélez & Gweon, 2021) or incorrect (R. Parikh, 1994) information about the listener’s decision problem. If they produce an irrelevant utterance, the listener might infer the lack of information (Jara-Ettinger & Rubio-Fernandez, 2021) and initiate a repair (Clark & Marshall, 1981).

To see how such imperfect information might play out in real-world discourse, one could imagine two friends finishing dinner at a restaurant (inspired by Grice’s “out of petrol” example):

\begin{quote}
(A): There’s an ice cream shop round the corner.
(B): Ah, I’m lactose intolerant.
(A): Well, they have great sorbet!
\end{quote}

What is happening here? Intuitively, the real-world context creates an implicit QUD (Benz & Jasinskaja, 2017; van Kuppevelt, 1995): “What should we do next?” A’s first comment implies a
desire to get dessert together. However, their proposal suggests they lack knowledge of B’s food allergies (formally, A’s knowledge of the true world state $w$ is incomplete, leading to an incorrect model of B’s payoffs). B repairs this by stating their dietary limitations. A responds by updating B’s beliefs about the decision problem (formally, expanding the set of actions $A$ at the ice cream shop beyond typical dairy-based options). Such an exchange leverages expectations of decision-theoretic relevance (Benz & Van Rooij, 2007) in conjunction with prior knowledge about the world (Schank & Abelson, 1977) to construct a mutually acceptable plan (Clark, 1996).

Lastly, due to cognitive limitations, speakers may not fully reason through the listener’s decision-making process. Indeed, results from Experiment 1 suggest such reasoning may be cognitively expensive: optimizing for decision-theoretic utility correlated with longer response times (Table C2). Speakers also displayed a bias towards positive-valued utterances and away from negative ones (Figs. C5, C6). Positive-valued utterances may be preferred because they direct the listener towards the optimal action, rather than away from suboptimal ones. A boundedly-rational (Hawkins et al., 2021; Lieder & Griffiths, 2020; Simon, 1957) formulation of relevance could potentially capture this effect by incorporating the listener’s planning complexity into utterance costs (Jara-Ettinger & Rubio-Fernandez, 2021b). This would bias the speaker towards utterances that produce relatively simple decision policies.

Our formulation of decision-theoretic relevance is undeniably simplistic relative to the complexity and uncertainty inherent in real-world communication. However, we are optimistic that can serve as a foundation for future work. We next discuss extensions to more nuanced utility functions and applications to more complex decision problems.

**Extensions to the speaker’s utility function**

Combining belief- and action-based objectives opens several important research directions. We first discuss our present model of truthfulness and relevance, then suggest extensions investigating the relationships between other belief- and action-based objectives.

The present work used a simple linear weighting of truthfulness and relevance (Eq. 9). However, this weighting is likely nonlinear and context-dependent. Our experiment tested a single type of decision context in a particular online population; future work should explore how
situational and cultural factors may affect it (Gibbs Jr & Colston, 2020; Mühlenbernd & Solt, 2022). We hypothesized that speaker uncertainty could be one such factor, with uncertainty over the listener’s decision context increasing truthfulness. Our pre-registered experiment failed to find an effect (Appendix E), but we believe exploring such context dependence is an important line of future work. For example, speakers may place greater value on truthfulness if they expect to interact with the listener again, or when describing more permanent aspects of the world versus more transient ones. Indeed, telling the time (Gibbs Jr & Bryant, 2008; van der Henst et al., 2002) may be a relatively extreme situation: the time is transient but important for planning actions. In such a domain, speakers may rightfully emphasize relevance over literal truthfulness.

Our formulation of truthfulness was also simplistic: we used a symmetric epistemic utility (Eq. 5). However, corroborating results from the literature on deception (van Swol et al., 2012), Experiment 2 suggests that this function may be asymmetric, with a greater penalty placed on outright falsehoods than being under-informative by remaining silent. Exploring the precise form of the truthfulness utility function—or potentially deriving it from decision-theoretic relevance, by implicitly considering future decisions—is an important direction for future work.

Another direction could consider mixed-motive settings, relaxing the assumption that speaker and listener share a reward function (Cao et al., 2018; Franke et al., 2012; Jaques et al., 2019; Noukhovitch et al., 2021; Ostrom et al., 1992; Wagner, 2015). Various epistemic goals may then be derived from differing preferences over real-world actions. Persuasion (Barnett et al., 2022; Mercier & Sperber, 2011) or deception (Oey et al., 2022) could emerge as short-term strategies to achieve speaker-serving actions, while truthfulness (Sbardolini, 2022) or politeness (P. Brown & Levinson, 1987) could signal aligned interests (Yoon et al., 2018; Yoon et al., 2020) to preserve long-term cooperation (Baxter, 1984). Finally, humans use language to enforce social norms (Li & Tomasello, 2021; Vaish et al., 2011) and commitments (Kanngiesser et al., 2017; Ostrom et al., 1992). Norms and commitments could both be modeled as socially-induced utility functions over actions (Fig. 1C), capturing preferences over behaviors beyond intrinsic rewards.

Returning to the “ice cream shop” example, one could imagine it taking place on a first date instead. Recast in this light, the participants’ motives are less certain: B might be lying about having a lactose intolerance in an attempt to end the evening early while saving face. Unfortunately, A is not taking the hint.
Extensions to more complex decision settings

Building on classic signaling games (Lewis, 1969), our experiments used a single-context, fully-observable decision problem (Lattimore & Szepesvari, 2020). More complex settings could help explain different forms of language: for example, Sumers et al. (2022) used a multiple-context setting to compare the use of imperative and informative language. Communication in sequential (Puterman, 1994) or partially-observable (Kaelbling et al., 1998) decision problems could explain transmission of procedural knowledge (McCarthy et al., 2021; Thompson et al., 2022) or task representations (Ho, Abel, et al., 2022).

Sequential decision settings would also allow interleaved communication and action (Khani et al., 2018). Speakers could infer the listener’s belief states from their actions (Baker et al., 2017; Jara-Ettinger et al., 2016), relaxing the assumption of perfect speaker knowledge. Finally, a fully embodied model of joint action (Clark, 1996) would allow agents to act physically or communicatively in service of a shared goal (Fisac et al., 2020; Hadfield-Menell et al., 2016).

Conclusion

Taken together, our work establishes and evidences a new theoretical framework encompassing both epistemic and decision-theoretic utilities. Reconciling these divergent perspectives on communication allowed us to model the Gricean maxims of truthfulness and relevance as independent and equal objectives. In doing so, we aim to fulfill Grice’s longstanding desire to generalize his theory by expanding the purpose of communication from an exchange of information to “allow for such general purposes as influencing or directing the actions of others” (Grice, 1975, p. 47).
References


*Psychological review.*


https://doi.org/10.1109/JPROC.2012.2225812
Appendix A

The formal relationship between QUDs and decision problems.

Let’s begin by working out the decision-theoretic relevance (Eq. 8) for the case with exactly two world states and two actions:

\[
\begin{bmatrix}
R(a, w) & R(\neg a, w) \\
R(a, \neg w) & R(\neg a, \neg w)
\end{bmatrix} = \begin{bmatrix}
r \\
s
\end{bmatrix}
\]

The relevance of an utterance is given by:

\[
U_{\text{Relevance}}(u | w) = \sum_{a \in A} \pi_L(a | u) R(a, w)
\]

\[
= \pi_L(a | u) R(a, w) + \pi_L(\neg a | u) R(\neg a, w)
\]

\[
= r \cdot \frac{\exp\{\beta_L \sum W R(a, w) P_L(w | u)\}}{\sum_A \exp\{\beta_L \sum W R(a, w) P_L(w | u)\}} + s \cdot \frac{\exp\{\beta_L \sum W R(\neg a, w) P_L(w | u)\}}{\sum_A \exp\{\beta_L \sum W R(\neg a, w) P_L(w | u)\}}
\]

\[
= \frac{1}{Z} [r \cdot \exp\{\beta_L (s + (r - s) P_L(w | u))\} + s \cdot \exp\{\beta_L (r + (s - r) P_L(w | u))\}]
\]

where the normalization constant is given by

\[
Z = \exp\{\beta_L [s + (r - s) P_L(w | u)]\} + \exp\{\beta_L [r + (s - r) P_L(w | u)]\}
\]

Now, intuitively, suppose there’s a unique action in each world with arbitrarily high reward \(r\). Assume further that taking the wrong action is arbitrarily costly, \(s < r\). Then if we know the true world is \(w\), we should expect the decision-theoretic utility of an utterance to reduce to some function of epistemic utility, since the listener’s belief about which world they are in maps one-to-one to which action they will take. We begin by showing that there does, in fact, exist a special case of a decision problem where an exact equivalence holds. We will call this case the “identity” decision problem.

**Theorem 1.** There exists an “identity” decision problem \(\mathcal{D}_0\) that is equivalent to the epistemic objective.

**Proof.** Note that the utility only depends upon the ratio of \(s\) and \(r\), not their absolute values. Then without loss of generality, we may fix \(r = 0\) and \(s < 0\) such that the utility simplifies to:
In other words, we observe that $U_{\text{Relevance}}(u|w)$ takes the functional form of a logistic function on the listener’s beliefs, with asymptote $s$, slope $s - \beta_L$, and midpoint $P_L(w|u) = 0.5$.

Now, we will show that there exists some $s < 0$ such that $U_{\text{Relevance}}(u|w, A) = \ln P_L(w|u)$.

To declutter the calculation, we will set $\beta_L = 1$ and let $p = P_L(w|u)$ and define

$A(p) = W_0(p^{1-2p}(2p - 1) \ln p)$ where $W_0$ is the principal branch of the Lambert W-function (i.e. the unique function such that $Ae^A = p^{1-2p}(2p - 1) \ln p$). Then consider the decision problem given by

$$s = \frac{\ln[A/((2p - 1) \ln p)]}{1 - 2p}$$

(A1)

yielding the utility

$$U_{\text{Relevance}}(u|w) = \frac{s}{1 + \exp{s[1 - 2p]}}$$

$$= \frac{\ln A - \ln[(2p - 1) \ln p]}{(1 - 2p)(1 + \frac{A}{(2p - 1) \ln p})}.$$ 

By definition of the W-function, we have

$$Ae^A = p^{1-2p}(2p - 1) \ln p$$

$\Rightarrow$ \hspace{0.5cm} $\ln A + A = \ln p^{1-2p} + \ln[(2p - 1) \ln p]$ 

$\Rightarrow$ \hspace{0.5cm} $\ln A - \ln[(2p - 1) \ln p] = (1 - 2p) \ln p - A$

Then, if we substitute the right side into the numerator of the utility, we obtain

$$U_{\text{Relevance}}(u|w) = \frac{(1 - 2p) \ln p - A}{(1 - 2p)(1 + \frac{A}{(2p - 1) \ln p})}$$

$$= \frac{(1 - 2p)(\ln p + \frac{A}{2p - 1})}{(1 - 2p)(1 + \frac{A}{(2p - 1) \ln p})}$$

$$= \frac{\ln p \cdot (1 + \frac{A}{(2p - 1) \ln p})}{1 + \frac{A}{(2p - 1) \ln p}}$$

$$= \ln p$$
obtaining the standard information-theoretic objective.

It only remains to check whether $s$ as defined in Eq. [A1] is real-valued for all $p \in [0,1]$ (i.e. that it does not have any singularities in the relevant domain). Our first concern is whether $A(p) = W_0(p^{1-2p}(2p-1)\ln p)$ even exists over $p \in [0,1]$ or whether it vanishes or blows up at some point. Note that the principal branch of the Lambert $W$-function is known to have real-valued solutions for $x \neq -1/e$, and it can be shown by taking the derivative that

$$
\min\{p^{1-2p}(2p-1)\ln p\} \approx -0.17 > -0.36 \approx -1/e \quad \text{(where the minimum is obtained at } p = e^{W_0(e/2)-1} \approx 0.73). \text{ Thus, } A \text{ itself is well-behaved over } p \in [0,1].
$$

Now we are ready to consider the full function $s = f(p)$.

An important consequence of this theorem is that the rewards $r$ and $s$ associated with different actions under the identity decision problem are determined purely as a function of the listener’s updated beliefs about the world after hearing the utterance, which are determined \textit{a priori} by the semantics of the language without reference to actions (otherwise any definition of an identity decision problem would be circular).

Given our definition of the identity decision problem, we are ready to consider the relationship to the question-under-discussion (QUD) framework proposed by Roberts (2012). Intuitively, the primary formal innovation of our model can be viewed as a way of relaxing the ‘hard’ Boolean notion of a QUD partition to a more graded notion derived from decision-theoretic principles. Rather than assigning each state to a cell of a partition, we assign each state a continuous value. In this section, we consider a basic theorem that sketches out the formal relationship between these constructs. This theorem says that grounding relevance in continuous decision problems is a strict generalization of the discrete partitions used as QUDs: we lose none of the formal expressiveness of QUDs, as any QUD can be translated to an equivalent decision
problem (specifically, a generalization of the identity decision problem from Theorem 1).

**Theorem 2.** Any partition-based QUD is equivalent to some decision problem.

**Proof.** Let $\sim_Q$ be an equivalence relation on possible worlds $w \in W$. This relation defines a canonical projection $f_Q : w \rightarrow [w]_Q$, where $[w]_Q = \{w' \in W : w' \sim_Q w\}$ is the equivalence class of an element $w$. Note that the set of equivalence classes form a partition on $W$. Under the RSA formulation (e.g. Kao, Wu, et al., 2014), a speaker aiming to be relevant with respect to $Q$ uses the following utility:

$$U_{QUD}(u; w^*, Q) = \ln P_L([w^*]_Q | u) \quad (A2)$$

where $w^*$ is the (known) true world state and $P_L$ is the probability assigned to the cell of the partition containing the true world:

$$P_L([w^*]_Q | u) = \sum_{w \in [w^*]_Q} P_L(w | u) \quad (A3)$$

Now, consider the following extension of the identity decision problem:

$$R(a_1, w) = \begin{cases} r & \text{if } w \in [w^*]_Q \\ s & \text{o.w.} \end{cases}$$

$$R(a_2, w) = \begin{cases} s & \text{if } w \in [w^*]_Q \\ r & \text{o.w.} \end{cases}$$

where the listener obtains reward $r$ upon taking action $a_1$ in the cell of the partition containing the true world and $a_2$ in any other cell. Then

$$U_{Relevance}(u | w^*) = \sum_{a \in A} \pi_L(a | u) R(a, w^*)$$

$$= \pi_L(a_1 | u) R(a_1, w^*) + \pi_L(a_2 | u) R(a_2, w^*)$$

$$= \frac{r \exp\{\beta_L \sum_w R(a_1, w) P_L(w | u)\}}{\sum_a \exp\{\beta_L \sum_w R(a, w) P_L(w | u)\}} + \frac{s \exp\{\beta_L \sum_w R(a_2, w) P_L(w | u)\}}{\sum_a \exp\{\beta_L \sum_w R(a, w) P_L(w | u)\}}$$

$$= \frac{1}{Z} \left[ r \exp\{\beta_L r P_L([w]_Q | u) + s P_L([-w]_Q | u) \} + s \exp\{\beta_L r P_L([-w]_Q | u) + s P_L([w]_Q | u) \} \right]$$
The rest of the proof follows the same calculation used in Theorem 1; we show that under the identity decision problem, the decision-theoretic utility given by Eq. 8 is equivalent to the epistemic utility under the QUD, given by Eq. A.2.

While the decision-theoretic formulation may be more theoretically satisfying for problems that arise in cognitive science, it can also be seen that the decision-theoretic approach is equivalent to an extension of the traditional set-theoretic formulation known as a value-weighted partition, where a real number is associated with each cell and a corresponding term is added to the speaker utility. Thus, the converse does not necessarily hold: there exist decision problems that cannot be expressed as partition-based QUDs.
Appendix B

Simulation Parameter Sensitivity

Our simulations of existing theoretical puzzles are intended to provide insight into how decision-theoretic relevance affects speakers’ choice of utterance and listeners’ inferences about their meanings. Our approach was to write down simple and intuitive payoff matrices, then use a gridsearch to identify plausible effect sizes in each case. In this section, we include the results of the grid searches to provide intuition about the sensitivity of the effects of interest to model parameters. In general, the effect for the first simulation (Grice’s “Out of Petrol” example) holds across a wide range of parameter space, while the second two simulations (the time-based examples) require relatively high $\lambda$ values.

It is important to emphasize that these simulations should be taken qualitatively rather than qualitatively. They illustrate that the combined model is necessary to drive intuitive effects, but the examples have many degrees of freedom (e.g., payoff matrices and utterance costs) and are not intended as evidence about actual parameter values. For example, incorporating utterance costs into all of the examples would expand the regions of parameter space in which the theoretically desired effects are found.
Figure B1

*Parameter sensitivity in Grice's “Out of Petrol” example.*

*Note.* Heatmap showing the listener’s posterior probability that the garage is open (as opposed to closed or nonexistent). The theoretically important contrast is between the “Truth-only” model ($\lambda = 0$) and the “Combined” model ($0 > \lambda > 1$); the “truth-only” model can never derive the implicature that the garage is likely to be open. When the speaker is assumed to be noisy (low $\beta$) or epistemically unreliable ($\lambda \gg .5$), the listener maintains some probability that the gas station simply does not exist.

Figure B2

*Parameter sensitivity in the “Lecture Start Time” example.*

*Note.* Heatmap showing the listener’s posterior probability that the lecture will start late. “Pragmatic loosening” occurs anytime the listener’s posterior is nonzero; this effect requires a relatively high $\lambda$ parameter. Incorporating utterance costs as suggested by Wilson and Sperber (2002b) would expand the range of parameter space in which the effect size is found.
Figure B3

Parameter sensitivity in the “Telling the Time” example.

Note. Line plot showing the speaker’s probability of rounding as a function of context and the \( \lambda \) parameter, assuming \( \beta_S = \beta_L = 10 \). Solid lines show the model predictions and dashed lines show empirically observed rates of rounding. \( \lambda > .5 \) is required to obtain rounding effects, and the empirical rounding behavior in both conditions is found when \( \lambda = .9 \).
Appendix C

Experiment 1

Methods Details

Feature Randomizations

To mitigate perceptual biases, the feature-value mappings were randomized for each participant. One set of features (color or texture) was assigned values in \([-1, 0, 1]\) without replacement, while the other was assigned values from \([-2, 0, 2]\) without replacement. Fig. C1A shows the “canonical” feature-value mapping, while Fig. C1B shows one of the randomized feature-value mappings seen by participants. For analysis purposes, we present all results mapped back to the “canonical” features. The ordering of trials was randomized. The order of actions in each trial (e.g. left-center-right) was randomized. To avoid response bias, the ordering of features was randomized (keeping textures and colors grouped); the ordering of values was randomized between ascending and descending.

Figure C1

Signaling bandits cast in a mushroom foraging setting. A: the “canonical” feature mapping, equivalent to Fig. 6A and used for analysis throughout. B: an example of a randomized feature mapping used in the experiment. At any point during the main experiment (after passing the quiz), participants could access this view by clicking on the “View Mushroom Info” button visible in other screenshots.
Attention Checks

In addition to the 28 experimental trials, we included eight attention checks which were identical across all participants. These trials followed the same basic structure as the experimental trials: a tourist was shown visiting a patch of three mushrooms. However, in these trials, the participants’ choice of utterance was constrained to a single feature and two possible values. The contexts and features were selected to ensure that truthfulness and relevance objectives were aligned: the true utterance (e.g. “Spots are +1”) always had positive decision-theoretic utility, while the false option (e.g. “Spots are -1”) had negative decision-theoretic utility. A cooperative speaker should always choose the true-and-useful message. Participants had to select the correct answer on at least 75% (6/8) of the attention checks. 273 of 285 participants (96%) passed the attention checks; the remaining 12 were still paid the $2 completion bonus but were excluded from the analysis.

Results Details

Manipulation Checks

We first obtain maximum likelihood estimates for each condition separately (Table C1). The ordering of the $\lambda$ parameter matches our expectations: $\lambda_{\text{Relevance-biased}} > \lambda_{\text{Unbiased}} > \lambda_{\text{Truth-biased}}$. To confirm that these results are significant, we perform a model comparison across conditions. We obtain marginal likelihoods for three models: (1) a single $\lambda$ parameter for all participants across all conditions; (2) independent $\lambda$ parameters for each condition, (3) ordinal $\lambda$ parameters for each condition, restricted to our hypothesis that $\lambda_{\text{Relevance-biased}} > \lambda_{\text{Unbiased}} > \lambda_{\text{Truth-biased}}$. We use a gridsearch over the same range of $\beta$ parameters: $\beta_S \in [1, 10]$ and $\beta_L \in [1, 10]$.

We find evidence in favor of the ordinal model in both cases: for ordinal-vs-single, the log BF = 1019.40; for the ordinal-vs-independent, the log BF = 2.09 (BF=8.07).

Likelihood Ratio Tests

Setting the $\lambda$ parameter (Eq. 9) to 0 yields a purely truthful speaker; setting it to 1 yields a pure relevance speaker; intermediate values yield a combined speaker. We use the
Figure C2

Example trial from Experiment 1.

Note. Left: participants were presented with a tourist visiting a particular mushroom patch (e.g. decision context). Right: participants selected utterances using drop-down menus.

Maximum-likelihood estimates obtained from gridsearch (Table C1) to perform likelihood ratio tests on the Unbiased condition. These confirm the results obtained by Bayes factors in the main text: the “Combined” model is significantly better than the “Relevance only” model ($\chi^2(1) = 2956.62, p < .0001$) and the “Truthfulness only” model ($\chi^2(2) = 3026.66, p < .0001$).

**Modeling Utterance Cost (Valence Bias)**

Model residuals (Fig. 8, C5) suggest that participants favored positive utterances (e.g., “Green, +2”) at higher rates than predicted by the model, and disfavored negative ones (e.g. “Blue, -2”). Intuitively, this may be explained by greater processing effort associated with
Figure C3

Instructions given to control conditions.

A Your Responsibilities as a Tour Guide

Your job is to teach tourists facts about mushroom features. It does not matter what mushrooms they choose. Always, and only, say true facts.

B Your Responsibilities as a Tour Guide

Your job is to ensure tourists choose tasty mushrooms. It does not matter if you tell the truth or not. You are allowed to lie.

Note. Conditions were biased towards the theoretical speaker models. A: Screenshot from “Truth-biased” instructions. B: Same for “Relevance-biased” instructions. “Unbiased” participants skipped this instruction page. See the experiment URL for the full instructions in context.

Negative-valence utterances: because they tell the listener to avoid certain actions, it requires an additional step to reason about which actions they are more likely to choose.

We tested a variant of the model including an additional cost term favoring positive-valence utterances (e.g. any utterance with a value of +1 or +2) and disfavoring negative-valence ones (-1 or -2):

\[ C(u) = \begin{cases} 
\nu & \text{if } u_R > 0 \\
-\nu & \text{if } u_R < 0 \\
0 & \text{if } u_R = 0 
\end{cases} \]  

(C1)

We used this term directly in Combined speaker (Eq. 9) and ran the same gridsearch as above, including the additional parameter \( \nu \in [0, .25, .5, .75, 1] \). We found that \( \nu = .25 \) improved the model fit and accounted for this residual structure (Fig. C6 C7). However, it did not substantially change the resulting maximum-likelihood parameters for the model. The full MLE for the Unbiased condition was \( \lambda = .55, \beta_S = 3, \beta_L = 2, \nu = .25 \).

Response Times
### Table C1

*Maximum likelihood estimates for different models and conditions in Experiment 1.*

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Model</th>
<th>Maximum Likelihood</th>
<th>λ</th>
<th>β₁</th>
<th>β₂</th>
<th>Log LIH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Truth Biased</td>
<td>87</td>
<td>Truth Only</td>
<td>-2 -</td>
<td>-5005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relevance Only</td>
<td>-2 1</td>
<td>-6890</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Combined</td>
<td>.35 3 1</td>
<td>-4556</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Truth Only</td>
<td>-2 -</td>
<td>-6570</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relevance Only</td>
<td>-2 2</td>
<td>-6535</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Combined</td>
<td>.55 3 3</td>
<td>-5057</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Relevance Biased</td>
<td>91</td>
<td>Truth Only</td>
<td>-2 -</td>
<td>-7346</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relevance Only</td>
<td>-3 2</td>
<td>-4768</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Combined</td>
<td>.85 4 2</td>
<td>-4403</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table C2

*Predicting per-trial response times.*

<table>
<thead>
<tr>
<th>Effect</th>
<th>Group</th>
<th>Term</th>
<th>Estimate</th>
<th>Std Error</th>
<th>Statistic</th>
<th>DOF</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Fixed</td>
<td>(Intercept)</td>
<td>10.73</td>
<td>0.56</td>
<td>19.27</td>
<td>271.00</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>2</td>
<td>Fixed</td>
<td>λ</td>
<td>2.99</td>
<td>0.90</td>
<td>3.32</td>
<td>271.00</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>3</td>
<td>Random</td>
<td>workerid</td>
<td>sd_Int</td>
<td>4.54</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Random</td>
<td>Residual</td>
<td>sd_Obs</td>
<td>10.13</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Linear mixed-effects model to predict per-trial response time with a fixed effect of the participant’s inferred λ parameter and random effects for each participant. λ correlates positively and significantly with response times.
Figure C4

Example comprehension questions.

Note. All participants took a 16-question comprehension quiz requiring them to learn the feature-value mapping, as well as important gameplay dynamics. The six questions regarding feature values are shown here. This ensured that all participants knew the true world state $w$. 
Figure C5

Model residuals.

Note. Model residuals (empirical - prediction) grouped by utterance value and feature. Participants displayed a bias for positive-valence features and utterances.

Figure C6

Modeling a preference for positive-valued utterances.

Note. Variance explained when incorporating an additional parameter favoring positive utterances and down-weighting negative ones.
Figure C7

Model residuals after incorporating utterance costs.

Note. Incorporating an additional parameter reflecting a bias towards positive-valence utterances substantially reduced residual structure (cf. Fig. C5 for residuals without this parameter).
Appendix D

Experiment 2

Methods Details

Generating Utterance-Context Pairs

To generate a set of utterances for Experiment 2, we first needed to assign each an decision-theoretic utility. We did so by simulating the behavior of a listener with $\beta_L = 3$ (we chose this value based on MLE estimates for “Unbiased” Exp 1 pilot data). This gave us 2,250 possible utterance-context pairs (30 utterances $\times$ 84 contexts).

Figure D1

Screenshots of actual user interface from Experiment 2.

Note. (A) an “aligned” utterance that is true (the “Spotted” feature is actually worth +1) and has positive decision-theoretic utility. (B) a “conflicted” utterance that is true but has negative decision-theoretic utility. (C) a “conflicted” utterance that is false but has positive decision-theoretic utility.

We next identified the theoretically interesting context-utterance pairs that put the objectives in “conflict” (false messages with positive decision-theoretic utility, or true messages with negative decision-theoretic utility). These consisted of utterances about “weak” features (whose true value was +1 or -1) or “neutral” features (whose true value was 0). We stratified these by (feature, value, decision-theoretic utility), and sampled one message from each. This yielded 72 “conflicting” context-utterance pairs with decision-theoretic utility in the range [-1.5, 2.3]. We balanced these using a similar procedure for “aligned” context-utterance pairs (i.e. true
messages with positive or zero decision-theoretic utility, or false messages with negative
decision-theoretic utility). We filtered for utterances in the decision-theoretic utility range [-1.5, 
2.3], stratified by (feature, value, decision-theoretic utility) and took one sample from each, 
yielding 94 pairs. We randomly sampled 72 of these, yielding a total of 144 utterances.

Finally, to ensure each participant saw a balanced set of utterances, we grouped them by 
feature, truthfulness and sorted them by decision-theoretic utility, then round-robin assigned 
them into four sets of 36. Participants were randomly assigned one of these sets. The ordering of 
trials was randomized.

**Attention Checks**

Attention checks follow the same structure as experimental trials. However there were two 
key differences: first, they were constant across all participants; and second, rather than a graded 
slider, participants were asked to make a binary choice between endorsing the utterance or stay 
silent. Attention check trials were selected to align the two speaker objectives: false utterances 
that were negative utility, and true utterances that were positive utility. As a result, cooperative 
speakers should always choose to say the true messages and not say the false ones.

**Results Details**

**Manipulation Checks**

We follow the same procedure for manipulation checks as Experiment 1. Table D1 shows 
the maximum-likelihood estimates for each model and each condition. We then use a model 
comparison to test whether the ordering of the $\lambda$ parameter matches our expectations:

$\lambda_{\text{Relevance-biased}} > \lambda_{\text{Unbiased}} > \lambda_{\text{Truth-biased}}$. We again obtain marginal likelihoods for a model with 
(1) a single $\lambda$ parameter, (2) $\lambda$ parameters for each condition, (3) ordinal $\lambda$ parameters for each 
condition $\lambda_{\text{Relevance-biased}} > \lambda_{\text{Unbiased}} > \lambda_{\text{Truth-biased}}$. We use a gridsearch over the same range of $\beta$
parameters: $\beta_S \in [1,10]$ and $\beta_L \in [1,10]$.

We find evidence in favor of the ordinal model in both cases: for ordinal-vs-single, the 
$log BF = 1801.49$; for the ordinal-vs-independent, the $log BF = 2.09$ ($BF=8.07$).
**Likelihood Ratio Tests**

We followed the same procedure as Experiment 1, using the maximum-likelihood estimates obtained from gridsearch (Table D1) to perform likelihood ratio tests on the “Unbiased” condition. These confirm the results obtained by Bayes factors in the main text: the “Combined” model is significantly better than the “Relevance Only” model ($\chi^2(1) = 556.36, p < .0001$) and the “Truthfulness only” model ($\chi^2(2) = 1701.03, p < .0001$).

**Table D1**

*Maximum likelihood estimates for Experiment 2.*

<table>
<thead>
<tr>
<th>Condition</th>
<th>N</th>
<th>Model</th>
<th>Maximum Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\lambda$ $\beta_S$ $\beta_L$ Log LH</td>
<td></td>
</tr>
<tr>
<td>Truth-biased</td>
<td>71</td>
<td>Truthfulness Only - 3 - -11832</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relevance Only  - 1 1 -14410</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Combined .15 3 1 -11813</td>
<td></td>
</tr>
<tr>
<td>Unbiased</td>
<td>78</td>
<td>Truthfulness Only - 1 1 -14962</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relevance Only  - 2 1 -14390</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Combined .75 3 1 -14112</td>
<td></td>
</tr>
<tr>
<td>Relevance-biased</td>
<td>79</td>
<td>Truthfulness Only - 1 - -15838</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Relevance Only  - 4 1 -13740</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Combined .90 3 2 -13709</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* The decision-theoretic utility of an utterance-context pair depends on the inferred $\beta_L$ parameter, as this determines how much the utterance will affect the listener’s policy. In order to use the same x-axis in all plots, Fig. 10 assumes $\beta_L = 1$. To generate the model predictions for the Relevance-biased condition, we use a different set of model parameters: $\lambda = .95, \beta_S = 4, \beta_L = 1$. This was the second-best parameter set from our grid search, with a negative log-likelihood of -13711.

**Testing for Bimodality**

The distribution of individually-estimated $\lambda$ parameters in the “Unbiased” condition appeared qualitatively bimodal (Fig. 10 bottom center), relative to the responses in Experiment
Table D2

*Model comparison for bimodality in Exp 2*

<table>
<thead>
<tr>
<th>Model</th>
<th>( \lambda )</th>
<th>Bayes factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined</td>
<td>( \lambda \sim \text{Uniform}(0,1) )</td>
<td>-</td>
</tr>
<tr>
<td>Bimodal</td>
<td>( \lambda \sim {0,1} )</td>
<td>(6.27 \times 10^{13})</td>
</tr>
</tbody>
</table>

1 which appeared uniform (Fig. 7, bottom center).

If responses in the “Unbiased” condition were sufficiently bimodal, the truthfulness-only and relevance-only models would provide the best explanation for individual participants’ response patterns: any individual could be modeled as *either* truth- or relevance-focused. Our finding that the “Combined” model is necessary to explain condition-level responses (Table 5) would be explained by population heterogeneity rather than individual participants blending the two utilities. In contrast, if the “Unbiased” condition was *not* strongly bimodal, this suggests that—even in our two-alternative forced-choice paradigm—*individuals* implement the “Combined” model and make graded tradeoffs between these two utilities.

To test for this bimodality, we fit two different condition-level models. Both allowed \( \lambda \) to vary by participant, but the *bimodal* model forced \( \lambda \sim \{0,1\} \) while the *combined* model allowed \( \lambda \sim \text{Uniform}(0,1) \). In each model, we fit a single \( \beta_S \) and \( \beta_L \) parameter for the population. The models were otherwise identical to those used in Experiment 2. Because the models now contain 80 parameters (78 participant-level \( \lambda \) parameters, \( \beta_S \) and \( \beta_L \)) we were not able to use the same grid search methodology to do a model comparison. We instead used annealed importance sampling (Grosse et al., 2016) to obtain likelihoods for each model. We ran MCMC chains with 100,000 samples, averaged the likelihood for each model across 50 such chains, and then used Bayes Factors to compare them (Table D2).

We found strong evidence in favor of the more flexible “Combined” model allowing \( \lambda \sim \text{Uniform}(0,1) \). Even in an experimental setting encouraging separation of the two speaker models, the “Combined” model still provides a better explanation for participant responses than the component models alone. *Individual* participants still made graded tradeoffs between truthfulness and relevance.
Appendix E

Experiment 3

In addition to the two experiments described in the main text, we pre-registered and ran a third experiment intended to test the hypothesis that uncertainty over the listener’s decision context would bias participants towards truthfulness. We followed the same endorsement paradigm and used the same trials as Experiment 2, and placed all participants in the “Unbiased” condition.

We then used a between participant manipulation to vary participants’ uncertainty in three conditions: “no” uncertainty, “moderate” uncertainty, and “high” uncertainty. Participants in the “no uncertainty” condition were told that tourists always visit a single mushroom patch, and—as in Experiments 1 and 2—each trial showed a patch with visible mushrooms. Participants in the “moderate uncertainty” condition were told that tourists always visit 2 patches (and take one mushroom from each). Each trial showed one known patch with visible mushrooms, and one unknown patch where the mushrooms are not visible. Finally, participants in the “high uncertainty” condition were told that tourists always visit 4 patches. Each trial showed one known patch and three unknown patches.

Our key prediction was that increasing uncertainty (i.e. more unknown patches) would make participants less willing to lie (i.e. the inferred $\lambda$ parameter would decrease as the uncertainty increased). However, our manipulation proved ineffective: participants in the “moderate” and “high” uncertainty conditions reported confusion about the “unknown” patches. All three conditions largely replicated the “Unbiased” condition from Experiment 2.

Methods Summary

We recruited 299 participants using Prolific using the same qualifications and compensation structure. 22 participants failed the comprehension check and 63 failed attention checks during the experiment, leaving a final sample size of 214. On average, participants spent about 16 minutes on the experiment ($M=16.01, SD=6.82$) and earned an hourly wage of $14.42.

Unfortunately, participant responses on the exit survey indicated confusion about the

\[19 \text{ Pre-registration: } \text{https://aspredicted.org/ZHC_MZQ} \]
presence of the additional mushroom patches in “Moderate” and “High” uncertainty conditions (Fig. E1), e.g. “I wasn’t really sure about what the mystery patch meant or what relevance that had.”

**Results Summary**

MLE estimates are provided in Table E1. Our pre-registered analysis followed a similar procedure as our manipulation checks in Experiment 1 and 2. We used a model comparison to test whether the ordering of the $\lambda$ parameter matches our predictions:

$\lambda_{\text{No Uncertainty}} > \lambda_{\text{Moderate Uncertainty}} > \lambda_{\text{High Uncertainty}}$. We again obtain marginal likelihoods for a model with (1) a single $\lambda$ parameter, (2) $\lambda$ parameters for each condition, (3) ordinal $\lambda$ parameters for each condition $\lambda_{\text{No Uncertainty}} > \lambda_{\text{Moderate Uncertainty}} > \lambda_{\text{High Uncertainty}}$. We use a gridsearch over the same range of $\beta$ parameters: $\beta_S \in [1,10]$ and $\beta_L \in [1,10]$. We find no evidence in favor of the ordinal model either case: for ordinal-vs-single, the BF $= 5.58e - 12$; for the ordinal-vs-independent, the BF $= 6.03e - 25$.

**Figure E1**

*Example trial in the “High Uncertainty” condition.*

*Note.* We hypothesized that the presence of additional “unknown” patches would bias participants towards truthfulness. However, participants reported confusion about the “unknown” patches and our results across conditions largely replicated Experiment 2’s Unbiased condition.
Table E1

Maximum likelihood estimates for Experiment 3.

<table>
<thead>
<tr>
<th>Uncertainty</th>
<th>N</th>
<th>Model</th>
<th>Maximum Likelihood</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>λ</td>
</tr>
<tr>
<td>“None”</td>
<td>74</td>
<td>Belief</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Action</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comb.</td>
<td>.70</td>
</tr>
<tr>
<td>“Moderate”</td>
<td>64</td>
<td>Belief</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Action</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comb.</td>
<td>.60</td>
</tr>
<tr>
<td>“High”</td>
<td>76</td>
<td>Belief</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Action</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comb.</td>
<td>.65</td>
</tr>
</tbody>
</table>

Note. Maximum likelihood estimates for different models and conditions in Experiment 3. Our “Uncertainty” manipulation had minimal effect on participant responses, likely due to reported confusion over the experimental user interface.
Figure E2

Results from Experiment 3.

Note. Top: response patterns across conditions. Scatterplots show individual responses, solid lines are nonparametric (locally-weighted) regressions summarizing the data, and dotted lines are MLE predictions. Participants displayed similar patterns across all three conditions. The most notable difference was a greater willingness to endorse true but negative decision-theoretic utterances under “Moderate” and “High” uncertainty (note that the yellow line’s Y-intercept is positive in both of these conditions). Bottom: MLE estimates for individual participants (histograms) and conditions (dashed lines and bolded text). Under “No” uncertainty (left), a substantial fraction of participants are nearly-pure “Relevance” speakers ($\lambda \approx 1$). Under “Moderate” and “High” uncertainty, fewer participants are pure “Relevance” speakers; however, this only results in a small shift in condition-level MLE estimates.