

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

It May or May Not Be Relevant: Testing the Relationship Between Relevance and Informativeness

Permalink

<https://escholarship.org/uc/item/7gz1k15k>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Authors

Feinmann, Diego

Ramotowska, Sonia

Publication Date

2026

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

It May or May Not Be Relevant: Testing the Relationship Between Relevance and Informativeness

Diego Feinmann (diego.feinmann@ipipan.waw.pl)¹ & Sonia Ramotowska^{2,3}

¹Institute of Computer Science Polish Academy of Sciences

²Institut Jean Nicod, Département d’Etudes Cognitives, ENS - PSL, EHESS, CNRS

³Institute for Logic, Language and Computation, University of Amsterdam

The authors contributed equally to this work; author order is alphabetical.

Abstract

The dominant view in cognitive science and formal pragmatics is that relevance is a matter of informativeness: a sentence is relevant if it is informative, and irrelevant otherwise. We provide experimental evidence against this thesis. For polar questions (e.g. *Is Mary coming to the party?*), we constructed two kinds of uninformative replies: replies that clearly violate relevance (uninformative-off-topic condition; e.g. *Mary lives in an old house*) and replies that are equally uninformative yet do not intuitively incur a relevance violation, such as uncertainty or disagreement reports (target condition; e.g. *Her parents disagree about that*). In a within-subject study ($N = 100$; 12 scenarios), participants judged both kinds of replies uninformative, but a forced-choice continuation diagnostic revealed a sharp dissociation in perceived relevance: target replies were judged relevant 95%, versus 3% for the uninformative-off-topic condition. These results show that informativeness is not a necessary condition for relevance.

Keywords: relevance; informativity; question answering; maxim of relation

The Notion of Relevance

Suppose you are a detective trying to determine the likelihood of the following hypothesis: «Jane died of natural causes». During your investigation, you encounter the following two pieces of information:

- (i) There were bloody footprints in the entrance hall of Jane’s apartment.
- (ii) Jane was a maths teacher.

Intuitively, (i) is relevant to the hypothesis at hand, whereas (ii) is not. What explains this difference? The dominant tradition in cognitive science and formal pragmatics—with early roots in Keynes (1921) and Carnap (1950)—ties relevance to *informativeness*: a sentence/proposition is relevant to a hypothesis just in case it makes a rational difference to one’s credence in that hypothesis. The familiar formal criterion is the following:

Def. I (INFO-RELEVANCE, CLASSICAL VERSION) Let Ω be the set of possible worlds, let P be a probability distribution over Ω , let ϕ be a sentence with truth-conditional content $[\phi] \subseteq \Omega$, and let $h \subseteq \Omega$ be a hypothesis. Then ϕ is **relevant** to h under P if and only if

$$P(h) \neq P(h \mid [\phi])^1$$

¹Def. I applies only when $P([\phi]) > 0$; otherwise $P(h \mid [\phi])$ is not defined.

In the detective example, P encodes the detective’s prior (her subjective probability assignments) while h is the proposition that Jane died of natural causes. If ϕ is instantiated as (i), then, given ordinary background assumptions, h is expected to become less likely, i.e. $P(h) > P(h \mid [(i)])$. Hence (i) is relevant to h . By contrast, if ϕ is instantiated as (ii), then h ’s probability is not expected to change, i.e. $P(h) = P(h \mid [(ii)])$ (for most plausible choices of P). In such a case, (ii) is irrelevant to h .

Relevance is not just for detectives, however: as Grice (1975), Lewis (1979), Sperber and Wilson (1986), and many others have argued, it plays a fundamental role in communication. This can be easy to miss, as overt relevance violations are relatively rare in ordinary conversation. For example, it would be highly unusual for us to ask *What time is it?* and be given *It’s raining!* as a response; and if that did happen, we would normally assume that our interlocutor misheard the question. Now suppose such an exchange really did take place—and that, after repeating the question more clearly, our interlocutor again replied *It’s raining!* Then we would face a problem: successful communication cannot proceed on that basis.

It is thus reasonable to ask what makes *It’s 3 pm!*, but not *It’s raining!*, relevant to the question *What time is it?* Fortunately, we need not look any further: the informativeness criterion implemented in Def. I can also be applied in such cases. To do so, however, a small adjustment is needed. Def. I evaluates relevance relative to a proposition (a subset of Ω), i.e. a single hypothesis. A question such as *What time is it?*, however, is not itself truth-evaluable and so cannot be identified with a single subset of Ω . The standard move, therefore, is to represent a question by the partition it induces on Ω (e.g. Groenendijk and Stokhof, 1984; Roberts, 1996/2012). This partition—known as the Question under Discussion (QUD)—is a set of mutually exclusive and jointly exhaustive hypotheses corresponding to the question’s possible complete answers. For example, in the case of *What time is it?*, each QUD cell corresponds to a possible time (or time interval). The updated definition is given below.

Def. II (INFO-RELEVANCE, QUD VERSION) Let P be a probability distribution over Ω , let Q be a QUD (a partition of Ω), and let ϕ be a sentence with truth-conditional content $[\phi] \subseteq \Omega$. Then ϕ is **relevant** to Q under P if and only if

$$\exists q \in Q \text{ such that } P(q) \neq P(q \mid [\phi]).$$

In words, Def. II can be paraphrased as follows: ϕ is relevant

to Q under P iff conditioning P on the truth of ϕ changes the probability of at least one cell of Q —that is, iff it shifts the probabilities assigned to the cells of Q (any gain for one cell must be offset by a loss for some other cell, and vice versa). Note that Def. II subsumes Def. I. In particular, in Def. II, the relevance of ϕ with respect to a single hypothesis h amounts to the relevance of ϕ with respect to the partition $\{h, \Omega \setminus h\}$. For this reason, we will take Def. II as our reference definition.

We also note that the standard informativeness-based definitions of relevance—the two definitions provided here—are somewhat idealised. In practice, exact probabilistic independence is rare: most pieces of information have *some* effect on most partitions, even if only a very slight one (e.g. learning that it is raining might slightly shift the probabilities assigned to the cells of the partition induced by *What time is it?*). For this reason, some accounts introduce a relevance threshold (see, e.g. Feinmann, 2020): a sentence counts as relevant to a QUD if and only if it has a sufficiently large (i.e. above-threshold) impact on the probabilities of the QUD cells. This issue will not concern us in the present study, since we will focus on cases where the probability shifts are either substantial or (at most) negligible.

To our knowledge, in virtually all Bayesian approaches to linguistic communication that are concerned with QUD relevance, Def. II is either adopted as the operative definition of relevance, or else the informativeness criterion it articulates (i.e. probability-shifting with respect to the QUD) is taken to be necessary for relevance (e.g. Büring, 2003; Feinmann, 2020; Hawkins and Goodman, 2017; Russell, 2012; van Rooy, 2004; see also Warstadt and Agha, 2022 for an experimental comparison of these approaches).² Indeed, across these accounts, a sentence/proposition counts as relevant only if it shifts the probabilities assigned to the QUD cells.³ In some of these accounts—for instance, van Rooy (2004)—relevance is subject to further constraints, but the informativeness criterion remains a necessary condition.

From Keynes (1921) to Feinmann (2026)

Is INFO-RELEVANCE on the right track? That is, does it capture human intuitions about what does—and does not—count as relevant?

Keynes (1921), long before Carnap (1950), considered this approach but ultimately rejected it. The reason for this rejection was conceptual in nature: a conjunctive sentence of the form $\lceil p \wedge q \rceil$, on Keynes' view, should count as relevant to a partition $\{h, \Omega \setminus h\}$ even if it leaves the probabilities of the

partition cells unchanged (i.e. $P(h \mid p \wedge q) = P(h)$) provided that conditioning on p alone and on q alone does affect those probabilities (i.e. $P(h \mid p) \neq P(h)$ and $P(h \mid q) \neq P(h)$).

In a recent article, Feinmann (2026) drew attention to contrasts such as the one in (1), which turns Keynes' conceptual intuition into an actual empirical challenge for INFO-RELEVANCE:

- (1) A: Is Mary coming to the party tonight?
 B₁: Mary's father thinks she is coming, while her mother thinks she isn't.
 B₂: Mary hates reading biographies.

The contrast is clear: (1B₁) seems like a relevant thing to say given the question raised, whereas (1B₂) does not. Yet, under reasonable assumptions (e.g. that Mary's mother is as reliable an informant as Mary's father), INFO-RELEVANCE predicts that neither (1B₁) nor (1B₂) should be perceived as a relevant contribution. Indeed, under reasonable assumptions, neither sentence is informative with respect to the QUD $\{\{\text{MARY IS COMING}\}, \{\text{MARY IS NOT COMING}\}\}$: conditioning on either sentence is expected to leave the probabilities of the QUD cells unchanged.

One could argue that in cases such as (1B₁) relevance is assessed incrementally as interpretation unfolds: upon hearing *Mary's father thinks she's coming*, the YES cell becomes more likely; upon hearing *her mother thinks she isn't*, it becomes less likely. If relevance can be assessed clause by clause in this way, the worry raised by (1B₁) may be mitigated: INFO-RELEVANCE would simply need to be understood as applying locally, rather than at the level of whole sentences.

The issue with this response, as Feinmann (2026) points out, is that, even when (1B₁) is replaced by a truth-functionally atomic paraphrase such as (2B₁), the contrast persists:

- (2) A: Is Mary coming to the party tonight?
 B₁: Mary's parents disagree about that.
 B₂: Mary hates reading biographies.

Unlike (1B₁), (2B₁) cannot be broken down in any obvious way into sub-clausal units that are independently informative relative to the QUD. Since the challenge posed by (2B₁) is, at least intuitively, the same as the one posed by (1B₁), the reply sketched above—namely, that INFO-RELEVANCE can be vindicated by incremental, clause-by-clause relevance assessment—does not seem to generalise.

Present Study

The aim of this study is to experimentally clarify the status of INFO-RELEVANCE, which entails that only sentences that are informative to the QUD count as relevant to it. To this end, we have tested Feinmann's cases—along with other potential counterexamples to INFO-RELEVANCE—in a controlled experiment.

The main difficulty in designing our experiment was methodological. The term 'relevant' has more than one meaning in natural language; as a result, if we ask participants

²In the psychology of reasoning literature, the link between relevance and informativeness is made explicit by Oaksford and Chater (1995) (see also Oaksford and Chater, 1994).

³Non-probabilistic approaches to relevance such as Groenendijk and Stokhof (1984) and Roberts (1996/2012), once lifted to a Bayesian setting, also entail this condition. On these accounts, a sentence is relevant to a QUD iff it eliminates at least one cell of the QUD (i.e. iff the truth-conditions of the uttered sentence fail to intersect at least one cell). In Bayesian terms, this amounts to assigning posterior probability 0 to at least one previously live cell.

whether a sentence is relevant given the question under consideration, their responses are likely to vary depending on how ‘relevant’ is interpreted.

To see this, consider the reply *Mary hates reading biographies* to the question *Is Mary coming to the party tonight?* It is perfectly natural to describe such a reply as *not relevant* (or as *irrelevant*). Here, ‘relevant’ means, roughly, *on-topic*, and ‘not relevant’ means *off-topic*.

Now consider a different use. Imagine a philosophy teacher explaining the notion of *validity*; she says: ‘We are checking whether this argument is valid—that is, whether the conclusion follows from the premises.’ A student raises her hand and objects, ‘But one of the premises is false.’ The teacher replies, ‘Yes, but that is not relevant.’ In this context, ‘not relevant’ does not mean off-topic—the student’s remark is perfectly on-topic. The issue is that it misses the point of the explanation (validity concerns whether the conclusion follows from the premises, not whether the premises are in fact true). These two possible construals of ‘not relevant’—off-topic vs. missing-the-point—are quite distinct, and an experimental prompt using the word ‘relevant’ risks conflating them.

Our goal was to elicit on-topic/off-topic judgments: whether an utterance is topically *related*, or *unrelated*, to the QUD. This is, we take it, the pre-theoretical notion of (ir)relevance that linguists are primarily interested in.⁴ However, ‘related’ is too vague a term to be used reliably in an experimental task. We also considered an explicit ‘on-topic/off-topic’ prompt, but worried that the wording might sound unnatural to some participants.

For these reasons, we adopted instead a linguistic diagnostic proposed by Feinmann (2026), which we take to provide a natural and robust way of eliciting the intended judgments of (ir)relevance. The idea is simple: in the event of a sentence that is irrelevant (in the relevant sense), it should always be felicitous to direct at it the following objection:

- (3) A: Is Mary coming to the party tonight?
B: Mary hates reading biographies.
A: *And what does that have to do with anything?*
- (4) A: Do you think it’s going to rain tomorrow?
B: Tomorrow is a bank holiday.
A: *And what does that have to do with anything?*

Conversely, such an objection is expected to be infelicitous when the response is perceived as a relevant contribution:

- (5) A: Is Mary coming to the party tonight?
B: I think she probably will.
A: *#And what does that have to do with anything?*

⁴Grice (1975), in a telling choice of label, called the relevance maxim the *Maxim of Relation*. Groenendijk and Stokhof (1984, p. 6), in turn, explicitly tied relevance to ‘what a conversation is about, to what the topic of a conversation is.’ Note also the following textbook gloss for the Maxim of Relation: ‘Stay on topic; don’t digress’ (Meyer, 2009, p. 56).

- (6) A: Do you think it’s going to rain tomorrow?
B: The weather forecast says it will not.
A: *#And what does that have to do with anything?*

This manipulation allows us to sidestep the problem of how to phrase the task prompt. If participants treat the follow-up *And what does that have to do with anything?* as appropriate, we take this to indicate that the target utterance is perceived as off-topic (and hence irrelevant in the intended sense). Conversely, if participants treat this follow-up as inappropriate, we take this to indicate that the target utterance is perceived as on-topic (and hence relevant). In short, the felicity of *And what does that have to do with anything?* serves as a diagnostic for perceived irrelevance, whereas its infelicity serves as a diagnostic for perceived relevance.

It is worth noting that INFO-RELEVANCE has no trouble accounting for the patterns in (3)–(6). Given a reasonable prior, it predicts B’s replies in (3) and (4) to be irrelevant—indeed, these responses are *not* expected to shift the probabilities of the QUD cells—and this is exactly what our diagnostic delivers, as reflected in the felicity of A’s retort. By contrast, INFO-RELEVANCE predicts B’s replies in (5) and (6) to be relevant—indeed, these responses shift the probabilities of the QUD cells—and our diagnostic again aligns with this prediction: A’s retort is infelicitous (#).

The crucial test for INFO-RELEVANCE, however, is whether *And what does that have to do with anything?* can be directed at responses such as (1B₁) and (2B₁). The intuition that Feinmann (2026) reports is that it cannot, which would place (1B₁) and (2B₁) in the same category as B’s replies in (5) and (6)—namely, in the relevant-contribution category. If this intuition can be experimentally validated, INFO-RELEVANCE is in trouble: this account treats uninformative sentences uniformly—all uninformative sentences are predicted to be irrelevant. Hence, if one group of uninformative sentences—for example, (1B₁) and (2B₁)—ends up being perceived as relevant while another ends up being perceived as irrelevant—for example, (3B) and (4B)—then INFO-RELEVANCE must be incorrect.

Methods

Participants

100 native English speakers were recruited on Prolific to participate in the experiment⁵. In addition to first language (English), participants were filtered by their approval rate (99–100%). Participants were paid £9/hour (pro-rated). The study was approved by the Faculty of Humanities Ethics Committee of the University of Amsterdam.

Materials and procedure

The experiment was implemented in Qualtrics. The design was within-subject. Each trial consisted of a scenario followed by two questions presented together on one screen—see

⁵The data for this experiment are available on OSF at <https://osf.io/d9n28/>.

Figure 1 for an example trial. The scenarios took the form of hypothetical dialogues between the participant performing the experiment and the fictional character Bob. The participant was given a QUD prompt—specifically, they were instructed to imagine that they asked Bob a polar question, e.g. whether Mary is coming to the party (see Table 1). Critical was Bob’s answer, which depended on the condition.

Imagine you and Bob are at a party. You ask Bob whether Mary is coming to the party.

Bob replies: **"Mary lives in an old house."**

Given what Bob just said, would you say that Mary is...

Less likely to come

Neither more likely nor less likely to come

More likely to come

Which of these two replies would you give to Bob? (*Choose the one that sounds more natural to you.*)

And what does that have to do with anything?

Alright, thanks.

Figure 1: Example trial.

Table 1: QUD prompts.

Trip	Imagine you and Bob are going on a trip. You ask Bob whether Jenny will be joining the trip.
Meeting	Imagine you and Bob are in a meeting. You ask Bob whether Steve will be attending the meeting too.
Party	Imagine you and Bob are at a party. You ask Bob whether Mary is coming to the party.
Thesis	Imagine you and Bob are talking about John. You ask Bob whether it is true that John has already finished his thesis.

In the INFORMATIVE-ON-TOPIC condition, Bob’s replies were straightforward answers to the polar QUD and, as such, were informative with respect to it and prima facie relevant/on-topic. We used two ‘positive’ answers (i.e. answers that shift probability mass towards the YES cell) and two ‘negative’ answers (i.e. answers that shift probability mass towards the NO cell). For each polarity (positive/negative), we included one fully resolving answer and one non-resolving, but strongly biasing, answer. See Items 9–12 in Table 2.

In the UNINFORMATIVE-OFF-TOPIC condition, the answers were uninformative—they did not provide any information as to which of the two QUD cells was more/less likely and, according to our intuitions, were irrelevant to the question. These are cases such as (QUESTION) *Is Mary coming to the*

party tonight?; (REPLY) *Mary hates reading biographies*. See Items 1–4 in Table 2.

The TARGET condition contained the critical stimuli: answers that, although intuitively uninformative with respect to the question, do not seem to incur a relevance violation. We selected four such cases; see Items 5–8 in Table 2. Items 7 and 8 were adaptations of examples discussed in Feinmann (2026), whereas the other two were new. We expected these answers not to be judged *informative*. Our goal was to determine whether they would nonetheless be judged *relevant*.

We used 12 items/scenarios in total, which were presented in random order. The first question participants had to answer was about the informativeness of Bob’s answer. Participants responded on a three-point scale indicating whether the probability of the QUD’s YES cell decreased, stayed the same, or increased after learning Bob’s answer. In this way, we measured whether the answer was judged to be informative.

Table 2: Item list for the experiment. Each item combines one of four QUD prompts (Trip, Meeting, Party, Thesis; see Table 1) with one of three response conditions: UNINFORMATIVE-OFF-TOPIC (UOT), TARGET (T), and INFORMATIVE-ON-TOPIC (IOT).

#	QUD	Bob’s answer	Condition
1	Trip	She thinks that she saw a famous actor yesterday.	UOT
2	Meeting	He has two brothers and one sister.	UOT
3	Party	Mary lives in an old house.	UOT
4	Thesis	His piano needs tuning.	UOT
5	Trip	She’ll decide at the last minute.	T
6	Meeting	He may or may not attend.	T
7	Party	Some people say she’ll come, but others say the opposite.	T
8	Thesis	John’s parents gave me conflicting reports about that.	T
9	Trip	She will.	IOT
10	Meeting	He won’t.	IOT
11	Party	I think she’ll probably come.	IOT
12	Thesis	I think he is unlikely to have finished.	IOT

The second question concerned the relevance of Bob’s answer. We measured relevance using a forced-choice measure based on Feinmann’s (2026) diagnostic. Participants were asked to choose which follow-up to Bob’s answer sounded more natural: *And what does that have to do with anything?* or *Alright, thanks* (presented in randomised order). Choosing the former was taken to indicate that Bob’s answer was perceived as irrelevant (i.e. calling for a relevance challenge), whereas choosing the latter was taken to indicate that it was perceived as relevant (*Alright, thanks* is a non-evaluative acknowledge-

ment; accordingly, selecting it—when the relevance challenge could have been chosen instead—signals that no relevance challenge is called for.)

Results

First, we tested whether participants' informativeness judgments, as measured by the first question, aligned with expectations. We recoded the responses "less likely" and "more likely" as *informative-answer* judgments, and "neither more likely nor less likely" as *uninformative-answer* judgments. Figure 2 shows the distribution of judgments in each experimental condition. The *informative-answer* judgments were dominant in the INFORMATIVE-ON-TOPIC, whereas the *uninformative-answer* judgments were dominant in the other two conditions.

The recorded binary variable was further used as the dependent variable in the mixed-effects logistic regression with condition (UNINFORMATIVE-OFF-TOPIC, INFORMATIVE-ON-TOPIC, TARGET) as a predictor. Random intercepts by-item and by-subject, and a by-subject random slope for condition were included. We used Helmert contrast coding to test (i) INFORMATIVE-ON-TOPIC versus the other two conditions and (ii) UNINFORMATIVE-OFF-TOPIC versus TARGET. The first contrast was significant ($\beta = 6.95, z = 8.25, p < 0.001$), indicating that answers in the INFORMATIVE-ON-TOPIC condition were much more likely to be judged informative than answers in the two other conditions. The second contrast was not significant ($\beta = -1.96, z = -1.78, p = 0.08$), indicating no reliable difference in the rate of *uninformative-answer* judgments between the UNINFORMATIVE-OFF-TOPIC and TARGET conditions. This is consistent with our expectation that answers in the TARGET condition should not be judged informative.

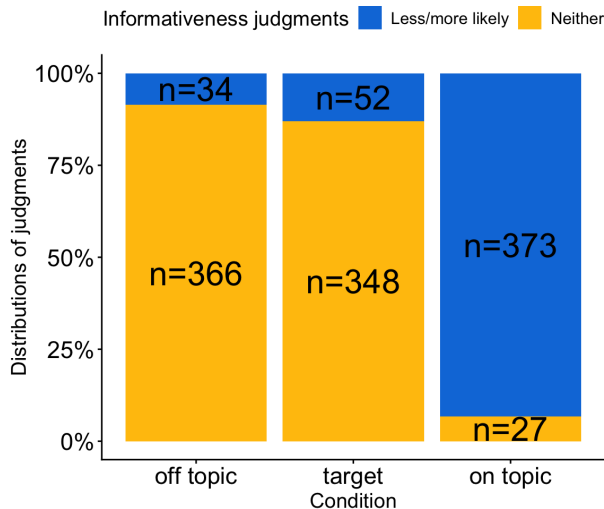


Figure 2: The distribution of informative and uninformative judgments (question 1) in the three experimental conditions (n indicates the number of each response type).

Next, we examined responses to the second question across

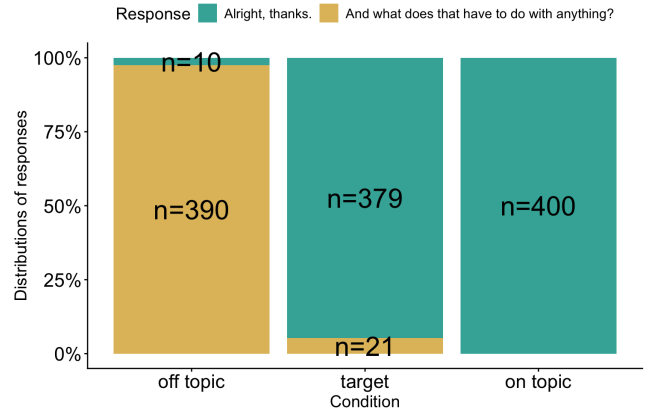


Figure 3: The distribution of relevance judgments (question 2) in the three experimental conditions (n indicates the number of each response type).

conditions (see Figure 3). We found that all participants judged the answers in the INFORMATIVE-ON-TOPIC condition as relevant (they all chose the *Alright, thanks* response). In the TARGET condition, 95% of judgments indicated that the answer was perceived as relevant, whereas in the UNINFORMATIVE-OFF-TOPIC condition, only 3% did. The mixed-effects logistic regression with response to the second question as dependent variable and condition as predictor (Helmert contrast coding as described above) revealed that the difference between the TARGET and the UNINFORMATIVE-OFF-TOPIC conditions was significant (second contrast: $\beta = 15.51, z = 7.4, p < 0.001$).⁶ This shows that, in the TARGET condition, replies were treated as relevant despite being judged uninformative (at rates not reliably different from UNINFORMATIVE-OFF-TOPIC).

Finally, we tested whether the informativeness judgments from the first question (binary encoded as described above) are a significant predictor of the relevance judgments from the second question. Contrary to the prediction of INFO-RELEVANCE, we found that informativeness judgments do not predict relevance judgments ($\chi^2(1) = 0.81, p = 0.37$).

Discussion

The main finding of this paper, we take it, is the following: relevance in the sense of being on-topic cannot be reduced to the informativeness criterion articulated in Def. II. In our task, replies from the TARGET condition were reliably judged uninformative—that is, participants did not in the main treat them as shifting probability mass towards either the YES or NO cell. Yet these same replies were overwhelmingly judged *relevant* by our relevance diagnostic, i.e. participants rejected the retort *And what does that have to do with anything?* and selected instead *Alright, thanks*. By contrast, the other set of uninformative replies—the replies in the UNINFORMATIVE-OFF-TOPIC condition—were overwhelmingly judged *irrelevant*

⁶The model included only by-subject random intercepts, as more complex models led to convergence problems.

by the same diagnostic. This selective dissociation is difficult to reconcile with INFO-RELEVANCE, according to which all replies that are uninformative should be perceived as irrelevant.

It is worth noting that the fact that informativeness fails to explain relevance does not mean that the notion is idle or unimportant. Consider the following experimental items (both used in our experiment):

- (7) [Item 7, Table 2]
Imagine that you and Bob are at a party. You ask Bob whether Mary is coming. Bob replies:
Some people say she'll come, but others say the opposite.
- (8) [Item 11, Table 2]
Imagine that you and Bob are at a party. You ask Bob whether Mary is coming. Bob replies:
I think she'll probably come.

As our results confirm, whereas Bob's replies in (7) and (8) are both perceived as relevant, only the reply in (8) can be regarded as informative. This difference has clear empirical consequences—for example, the reply in (7) can naturally be followed up with *Well, that's not very helpful...*, whereas the one in (8) cannot. Thus, even if informativeness fails to track perceived relevance, it does, tentatively, seem to track another important dimension of linguistic evaluation, namely *perceived helpfulness*.

Why might the uninformative replies from the TARGET condition be perceived as relevant? One natural hypothesis is that such replies convey meta-information about the conversational situation. Reports of disagreement (e.g. *Her parents disagree about that*) and explicit uncertainty (e.g. *He may or may not attend*) have no informational impact on the QUD, but they nevertheless respond to it in a discourse-pragmatic sense: they signal that the speaker is not in a position to answer, or that the available evidence is inconclusive, which may in turn guide the next conversational move (e.g. whether to seek further evidence or consult a different source). Of course, this is only an intuition: it is not yet clear to us how to turn this idea into a fully explicit theory of relevance, nor whether it generalises to all cases of perceived relevance without informativeness. Still, it suggests a promising direction for further work.

Looking ahead, an obvious question for further experimental work is what role, if any, informativeness should play in a theory of relevance. INFO-RELEVANCE effectively makes two claims: first, that replies which are uninformative with respect to the QUD should be perceived as irrelevant; and second, that replies which are informative with respect to the QUD should be perceived as relevant. Our task was designed to target the first claim, and the results suggest that it is false: some replies that are perceived as uninformative are nonetheless judged to be relevant. This is already enough to undermine the common assumption that QUD-level informativeness is a necessary condition for perceived relevance.

Still, the status of the second claim remains an open empirical question. Intuitively, it is plausible—and our data do not challenge it—but our materials were not designed to test it, since the informative replies we used were limited to complete answers (e.g. *She will (come)*) or replies structurally close to complete answers (e.g. *I think she'll probably come*). A natural next step would be to test a broader range of informative replies, including cases where the surface form diverges sharply from the question, e.g. for *Is Mary coming to the party tonight?*, replies such as *She has a difficult exam tomorrow* or *It's blowing a gale and the trains are cancelled*.

In the meantime, one conclusion seems hard to avoid: relevance judgments—at least as elicited by Feinmann (2026)'s diagnostic—cannot be reduced to QUD-level informativeness. Our contribution is to bring this point into sharper focus, in the hope that future work can extend our theoretical understanding of relevance.

Acknowledgments

For extremely valuable input at the early stages of this project, both authors are grateful to Jacopo Romoli. Sonia Ramotowska's position was funded by the "Nothing is Logical" project (the Dutch Research Council (NWO) OC grant 406.21.CTW.023), as well as the French National Research Agency (ANR) grants "Communicative efficiency, cognitive constraints and lexical meaning" (ANR-23-CE28-0016) and "FrontCog" (ANR-17-EURE-0017).

References

- Büring, D. (2003). On D-Trees, Beans, And B-Accents. *Linguistics and Philosophy*, 26(5), 511–545. <https://doi.org/10.1023/A:1025887707652>
- Carnap, R. (1950). *Logical foundations of probability*. University of Chicago Press.
- Feinmann, D. (2020). *Lexical illusions, (Non-)Maximality, and invisible gaps* [PhD Thesis]. University of Sheffield.
- Feinmann, D. (2026). A note on relevance. *Journal of Semantics*, ffaf013. <https://doi.org/10.1093/jos/ffaf013>
- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (pp. 41–58, Vol. 3). Academic Press.
- Groenendijk, J. A. G., & Stokhof, M. J. B. (1984). *Studies on the Semantics of Questions and the Pragmatics of Answers* [PhD Thesis]. University of Amsterdam.
- Hawkins, R. X. D., & Goodman, N. D. (2017). Why do you ask? The informational dynamics of questions and answers [PsyArXiv preprint]. <https://doi.org/10.31234/osf.io/j2cp6>
- Keynes, J. M. (1921). *Treatise on probability*. Macmillan & Co.
- Lewis, D. (1979). Scorekeeping in a language game. *Journal of Philosophical Logic*, 8(1), 339–359.
- Meyer, C. F. (2009). *Introducing English linguistics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511757822>

- Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, 101(4), 608–631. <https://doi.org/10.1037/0033-295X.101.4.608>
- Oaksford, M., & Chater, N. (1995). Information gain explains relevance which explains the selection task. *Cognition*, 57(1), 97–108. [https://doi.org/10.1016/0010-0277\(95\)00671-K](https://doi.org/10.1016/0010-0277(95)00671-K)
- Roberts, C. (1996/2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5(6), 1–69. <https://doi.org/10.3765/sp.5.6>
- Russell, B. (2012). *Probabilistic reasoning and the computation of scalar implicatures* [PhD thesis]. Brown University.
- Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Harvard University Press.
- van Rooy, R. (2004). Utility, informativity and protocols. *Journal of Philosophical Logic*, 33, 389–419. <https://doi.org/10.1023/B:LOGI.0000036830.62877.ee>
- Warstadt, A., & Agha, O. (2022). Testing Bayesian measures of relevance in discourse. *Proceedings of Sinn und Bedeutung* 26, 865–886. <https://doi.org/10.18148/sub/2022.v26i0.1034>