COMMUNICATION & COGNITIVE ARCHITECTURE

Week 5: Mindreading in Human Communication

Daniel W. Harris

Ch. 4 Outline:

1. I have posited a lot of mindreading

2. Some evidence that we mindread a lot, but that it is resource intensive

3. Graceful degradation vs. Progressive enhancement?

4. Some cost-saving proposals for communicative mindreading

3. Graceful degradation vs. Progressive enhancement?

- Nearly everyone agrees that we sometimes do fancy mindreading when we communicate.
- Nearly everyone agrees that we also sometimes do less fancy things.
- Is either of these ways of doing things more fundamental than the other?
- And in what sense?

Graceful degradation versus progressive enhancement

Main Page
Browse categories
Recent changes

Tools
What links here
Related changes
Special pages
Printable version
Permanent link
Page information

3. Graceful degradation vs. Progressive enhancement?

Graceful degradation in interface design
Providing an alternative version of your functionality or making the user aware of shortcomings of a product as a safety measure to ensure that the product is usable.

Progressive enhancement in interface design
Starting with a baseline of usable functionality, then increasing the richness of the user experience step by step by testing for support for enhancements before applying them.

3. Graceful degradation vs. Progressive enhancement?

Graceful degradation in communicative cognition
To use language in accordance with its proper function requires doing fancy planning and mindreading. We rely on less resource-intensive but less reliable backup options when that's not available.

Progressive enhancement communicative cognition
The most basic and central cases of language use are the ones without fancy planning and mindreading, which we wheel out only in unusual cases when something special is called for.

Minxin

...Westra & Nagel article makes me wonder the defeatability of the claim that mindreading takes a central role in human communications. In particular, since much of the theory relies on subconscious activities of the mind, there would not be any reliable introspective evidence or self-reported evidence that could serve as a potential defeater of the model. What kind of findings would it take for one to consider it improbable (or even impossible) that mind reading is not involved in most of our daily communications.

Minxin

For example, consider an attendant with a ridiculously huge grinder salting my steak. They stoped their motion of grinding the salt when they hear I say the word "when." Their action is a clear marker that they have understood my intention for them to stop salting, but prima farcie, my expression has nothing to do with stopping or enough or anything of the sort. But anyone who has learned the expression, either through someone is kind enough to teach them how to respond to "say when", or by a similar experience of humiliation to mine where I stunned and stared at a little Parmesan cheese hill being formed on top of my salad without knowing what to say, would know that my saying "when" is an indication for the attendant to stop grinding. The intention-recognition model hypothesizes that in reaching the conclusion that I should say "when," I formed the intention for the attendant to stop salting and planned this by expecting them to recognize my intention for this as I say "when."

If, immediately after I said "when," someone asked me "why did you say that?" I would probably give the explanation that I said that intending for the attendant to stop. Though this could serve as an indication that the intention-recognition model is true, it is also possible that this is merely an adhoc explanation of the fact. On the hand, if my initial response is "I don't know, just seemed like the thing to say when someone says 'say when.'" This would still not act as counter-evidence against the model as a supporter might just respond by claiming that I simply do not realize the unconscious communication process.

ANNUAL REVIEWS

Annual Review of Linguistics

The Rational Speech Act Framework

Judith Degen

Department of Linguistics, Stanford University, Stanford, California, USA;
email: jdegen@stanford.edu

Annu. Rev. Linguist. 2023. 9:26.1–26.22

The *Annual Review of Linguistics* is online at linguistics.annualreviews.org

https://doi.org/10.1146/annurev-linguistics-031220-010811

Keywords

probabilistic pragmatics, computational pragmatics, experimental pragmatics, experimental semantics, context

Abstract

The past decade has seen the rapid development of a new approach to pragmatics that attempts to integrate insights from formal and experimental semantics and pragmatics, psycholinguistics, and computational cognitive science in the study of meaning: probabilistic pragmatics. The most influential probabilistic approach to pragmatics is the Rational Speech Act (RSA) framework. In this review, I demonstrate the basic mechanics and commitments of RSA as well as some of its standard extensions, highlighting the key features that have led to its success in accounting for a wide variety of pragmatic phenomena. Fundamentally, it treats language as probabilistic, informativeness as gradient, alternatives as context-dependent, and subjective prior beliefs (world knowledge) as a crucial facet of interpretation. It also provides an integrated account of the link between production and interpretation. I highlight key challenges for RSA, which include scalability, the treatment of the boundedness of cognition, and the incremental and compositional nature of language.

Predicting Pragmatic Reasoning in Language Games

Michael C. Frank* and Noah D. Goodman

One of the most astonishing features of human language is its ability to convey information efficiently in context. Each utterance need not carry every detail; instead, listeners can infer speakers' intended meanings by assuming utterances convey only relevant information. These communicative inferences rely on the shared assumption that speakers are informative, but not more so than is necessary given the communicators' common knowledge and the task at hand. Many theories provide high-level accounts of these kinds of inferences (*1*–*3*), yet, perhaps because of the difficulty of formalizing notions like "informativeness" or "common knowledge," there have been few successes in making quantitative predictions about pragmatic inference in context.

We addressed this issue by studying simple referential communication games, like those described by Wittgenstein (*4*). Participants see a set of objects and are asked to bet which one is being referred to by a particular word. We modeled human behavior by assuming that a listener can use Bayesian inference to recover a speaker's intended referent r_S in context C, given that the speaker uttered word w:

$$P(r_s|w,C) = \frac{P(w|r_s,C)P(r_s)}{\sum_{r' \in C} P(w|r',C)P(r')} \quad (1)$$

This expression is the product of three terms: the prior probability $P(r_S)$ that an object would be referred to; the likelihood $P(w|r_S,C)$ that the speaker would utter a particular word to refer to the object; and the normalizing constant, a sum of these terms computed for all referents in the context.

We defined the prior probability of referring to an object as its contextual salience. This term picks out not just perceptually but also socially and conversationally salient objects, capturing the common knowledge that speaker and listener share, as it affects the communication game. Because there is no a priori method for computing this sort of salience, we instead measured it empirically (*5*).

Department of Psychology, Stanford University, Stanford, CA 94305, USA.

*To whom correspondence should be addressed. E-mail: mcfrank@stanford.edu

The likelihood term in our model is defined by the assumption that speakers choose words to be informative in context. We quantified the informativeness of a word by its surprisal, an information-theoretic measure of how much it reduces uncertainty about the referent. By assuming a rational actor model of the speaker, with utility defined in terms of surprisal, we can derive the regularity that speakers should choose words proportional to their specificity (*6*, *7*):

$$P(w|r_s,C) = \frac{|w|^{-1}}{\sum_{w' \in W} |w'|^{-1}} \quad (2)$$

where $|w|$ indicates the number of objects to which word w could apply and W indicates the set of words that apply to the speaker's intended referent.

In our experiment, three groups of participants each saw communicative contexts consisting of sets of objects varying on two dimensions (Fig. 1A). We systematically varied the distribution of features on these dimensions. To minimize the effects of particular configurations or features, we randomized all other aspects of the objects for each participant. The first group (speaker condition) bet on which word a speaker would use to describe a particular object, testing the likelihood portion of our model. The second group (salience condition) was told that a speaker had used an unknown word to refer to one of the objects and was asked to bet which object was being talked about, providing an empirical measure of the prior in our model. The third group (listener condition) was told that a speaker had used a single word (e.g., "blue") and again asked to bet on objects, testing the posterior predictions of our model.

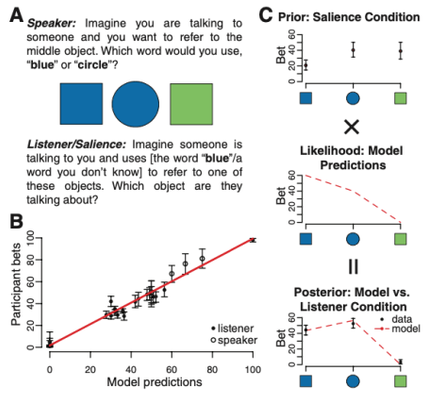


Fig. 1. (**A**) An example stimulus from our experiment, with instructions for speaker, listener, and salience conditions. (**B**) Human bets on the probability of a choosing a term (speaker condition, $N = 206$) or referring to an object (listener condition, $N = 263$), plotted by model predictions. Points represent mean bets for particular terms and objects for each context type. The red line shows the best linear fit to all data. (**C**) An example calculation in our model for the context type shown in (A). Empirical data from the salience condition constitute the prior term, $N = 20$ (top); this is multiplied by the model-derived likelihood term (middle). The resulting posterior model predictions (normalization step not shown) are plotted alongside human data from the listener condition, $N = 24$ (bottom). All error bars show 95% confidence intervals.

Mean bets in the speaker condition were highly correlated with our model's predictions for informative speakers ($r = 0.98$, $P < 0.001$; Fig. 1B, open circles). Judgments in the salience and listener conditions were not themselves correlated with one another ($r = 0.19$, $P = 0.40$), but when salience and informativeness terms were combined via our model, the result was highly correlated with listener judgments ($r = 0.99$, $P < 0.0001$, Fig. 1B, solid circles). This correlation remained highly significant when predictions of 0 and 100 were removed ($r = 0.87$, $P < 0.0001$). Figure 1C shows model calculations for one arrangement of objects.

Our simple model synthesizes and extends work on human communication from a number of different traditions, including early disambiguation models (*8*), game-theoretic signaling models (*9*), and systems for generating referring expressions (*10*). The combination of an information-theoretic definition of "informativeness" along with empirical measurements of common knowledge enables us to capture some of the richness of human pragmatic inference in context.

References and Notes

1. H. Grice, in *Syntax and Semantics*, P. Cole, J. Morgan, Eds. (Academic Press, New York, 1975), vol. 3, pp. 41–58.
2. D. Sperber, D. Wilson, *Relevance: Communication and Cognition* (Harvard Univ. Press, Cambridge, MA, 1986).
3. H. Clark, *Using Language* (Cambridge Univ. Press, Cambridge, 1996).
4. L. Wittgenstein, *Philosophical Investigations* (Blackwell, Oxford, 1953).
5. H. Clark, R. Schreuder, S. Buttrick, *J. Verbal Learn. Verbal Behav.* **22**, 245 (1983).
6. Materials and methods are available as supplementary materials on *Science* Online.
7. F. Xu, J. B. Tenenbaum, *Psychol. Rev.* **114**, 245 (2007).
8. S. Rosenberg, B. D. Cohen, *Science* **145**, 1201 (1964).
9. A. Benz, G. Jäger, R. Van Rooij, Eds., *Game Theory and Pragmatics* (Palgrave Macmillan, Hampshire, UK, 2005).
10. R. Dale, E. Reiter, *Cogn. Sci.* **19**, 233 (1995).

Supplementary Materials

www.sciencemag.org/cgi/content/full/336/6084/998/DC1
Materials and Methods
Supplementary Text

3 January 2012; accepted 10 April 2012
10.1126/science.1218633

Basic Scalar Implicature Game

Probability Space

The prior probabilities of each of the propositions in the meaning space.

$$P(m_0) = 0.2$$

$$P(m_1) = 0.2$$

$$P(m_2) = 0.2$$

$$P(m_3) = 0.2$$

$$P(m_4) = 0.2$$

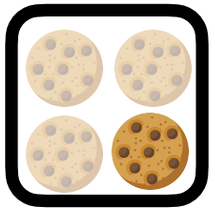
The Meaning Space

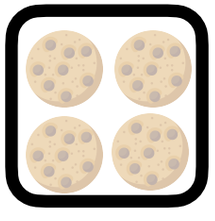
Things that S could have meant. Think of these as answers to the QUD: how many cookies did Alex eat?

m_0  All cookies remain.

m_1  One cookie eaten.

m_2  Two cookies eaten.

m_3  Three cookies eaten

m_4  All cookies eaten.

The Utterance Space

Things that S could have uttered.

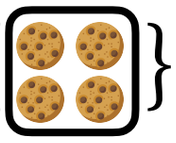
u_{none} = "Alex ate none of the cookies"

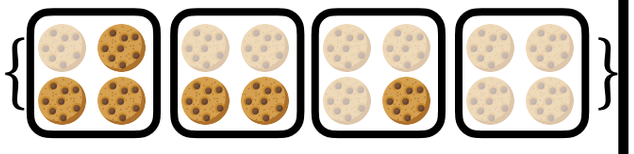
u_{some} = "Alex ate some of the cookies"

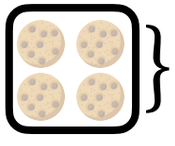
u_{all} = "Alex ate all of the cookies"

Semantics

The literal meanings of the utterances in U.

$[\![u_{none}]\!] = \{$  $\}$

$[\![u_{some}]\!] = \{$  $\}$

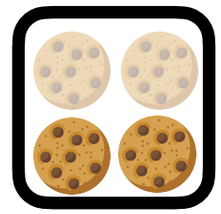
$[\![u_{all}]\!] = \{$  $\}$

Basic Scalar Implicature Game

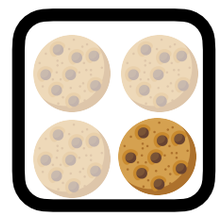
$\mathrm{P}(m_0) = 0.2$

$\mathrm{P}(m_1) = 0.2$

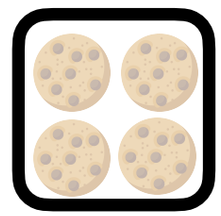
$\mathrm{P}(m_2) = 0.2$



$\mathrm{P}(m_3) = 0.2$

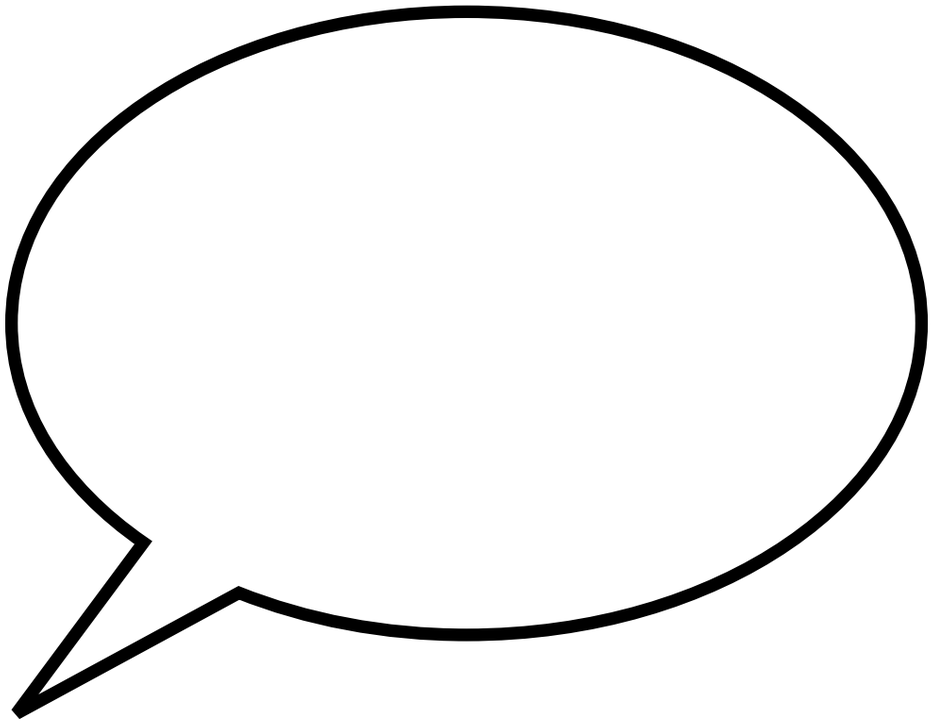


$\mathrm{P}(m_4) = 0.2$



Basic Scalar Implicature Game

- Suppose that S makes an utterance.
- L wants to know the meaning.
- Probabilistically: L's job is to infer the likelihood of each possible meaning, given that the speaker made that utterance.
- For each m, they calculate P(m|u): the probability of each meaning conditional on that utterance being made.

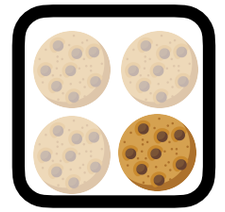


$P(m_0) = 0.2$

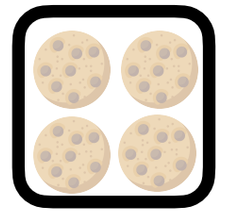
$P(m_1) = 0.2$

$P(m_2) = 0.2$

$P(m_3) = 0.2$



$P(m_4) = 0.2$



Basic Scalar Implicature Game

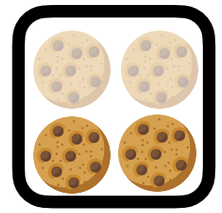
- The literal meaning of this utterance is incompatible with all of the meanings except m_4.
- So, if they assume that the speaker is knowledgeable, honest, and informative (i.e., the maxim of quality), they can infer the truth of m_4.

Alex ate all of the cookies.

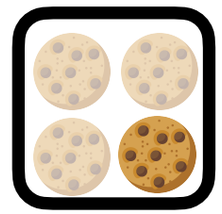
$$P(m_0|u_{all}) = 0$$

$$P(m_1|u_{all}) = 0$$

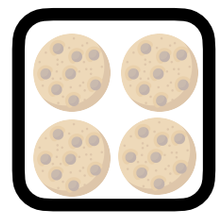
$$P(m_2|u_{all}) = 0$$



$$P(m_3|u_{all}) = 0$$



$$P(m_4|u_{all}) = 1$$



Basic Scalar Implicature Game

- Similar story here.

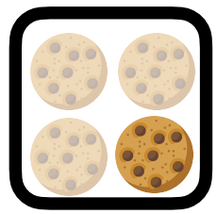
Alex ate none of the cookies.

$$P(m_0 | u_{none}) = 1$$

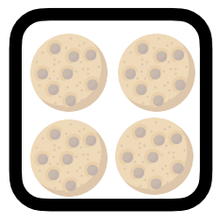
$$P(m_1 | u_{none}) = 0$$

$$P(m_2 | u_{none}) = 0$$

$$P(m_3 | u_{none}) = 0$$



$$P(m_4 | u_{none}) = 0$$



Basic Scalar Implicature Game

- What about when the speaker says this?
- On one hand, the literal meaning of this utterance only rules out one possibility, m_0.
- But in cases like this, we tend to detect a scalar implicature that also rules out (or lowers the probability of) m_4.
- How can we predict this?

Alex ate some of the cookies.

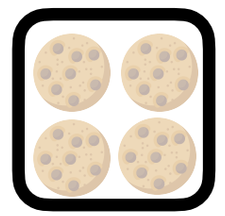
$\mathrm{P}(m_0) = 0.2$

$\mathrm{P}(m_1) = 0.2$

$\mathrm{P}(m_2) = 0.2$

$\mathrm{P}(m_3) = 0.2$

$\mathrm{P}(m_4) = 0.2$



Basic Scalar Implicature Game

- First, here's a rule that predicts what the literal listener (L0) would do:

$$P_{L0}(m|u) \propto \delta_{m \in [\![u]\!]} \cdot P(m)$$

$\delta_{m \in [\![u]\!]} = 1$ if m is one of u's meanings;
otherwise it is 0

- So L0 distributes probabilities across the literal meanings of u, in proportion to their prior probabilities.

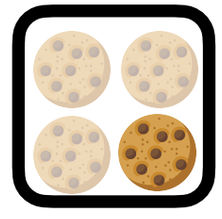
Alex ate some of the cookies.

$P_{L0}(m_0|u_{some}) = 0$

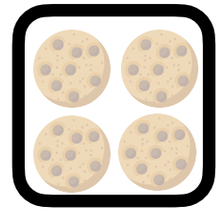
$P_{L0}(m_1|u_{some}) = .25$

$P_{L0}(m_2|u_{some}) = .25$

$P_{L0}(m_3|u_{some}) = .25$



$P_{L0}(m_4|u_{some}) = .25$



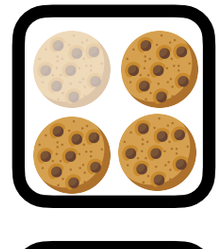
Basic Scalar Implicature Game

- In this case, since the priors were all even, we get the assumption that M_0 is ruled out, but M_1–M_4 are equally likely.
- This is the strictly literal interpretation.

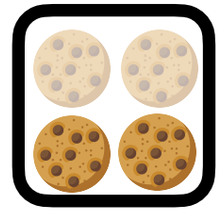
Alex ate some of the cookies.

$$P_{L0}(m_0|u_{some}) = 0$$

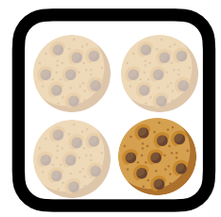
$$P_{L0}(m_1|u_{some}) = .25$$



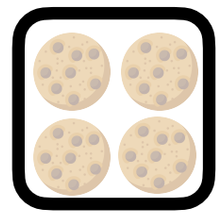
$$P_{L0}(m_2|u_{some}) = .25$$



$$P_{L0}(m_3|u_{some}) = .25$$

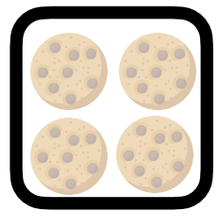


$$P_{L0}(m_4|u_{some}) = .25$$

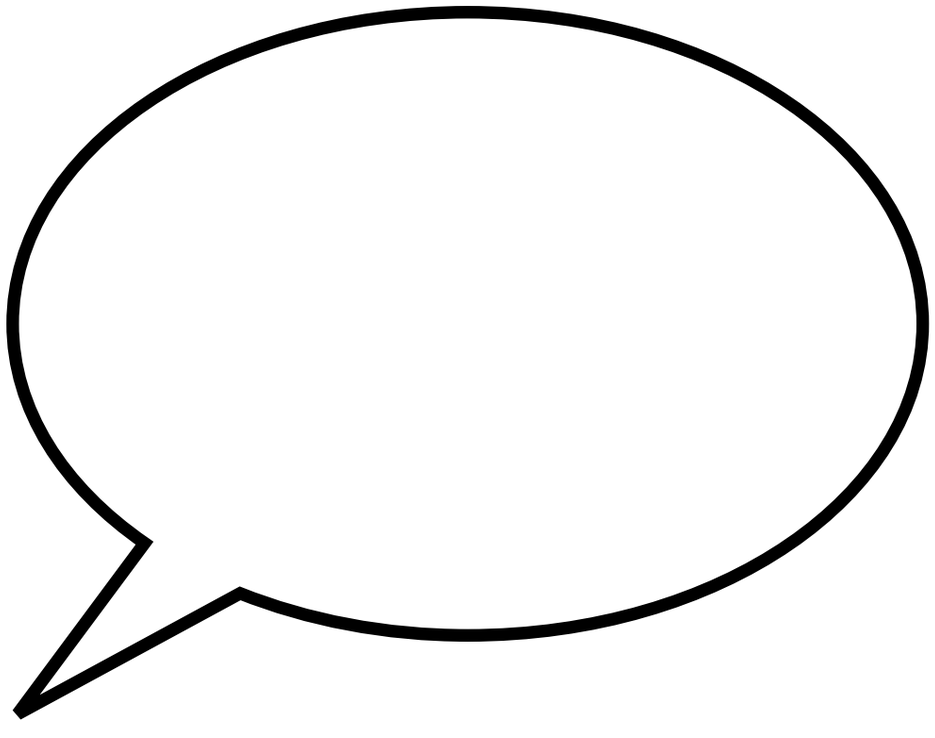


Basic Scalar Implicature Game

- Now let's think about this from the perspective of the "pragmatic speaker", S1—a speaker who is trying to be informative, and thinking about how the listener will interpret them.
- Given that they want to mean m, they need to calculate P(u|m) for each possible utterance u, which is the probability that they should utter u given how good it would be if the listener inferred m.



$$P(u_{\text{all}} | m_{1\text{-}3}) = ?$$



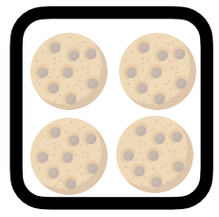
$$P(u_{\text{some}} | m_{1\text{-}3}) = ?$$

$$P(u_{\text{none}} | m_{1\text{-}3}) = ?$$

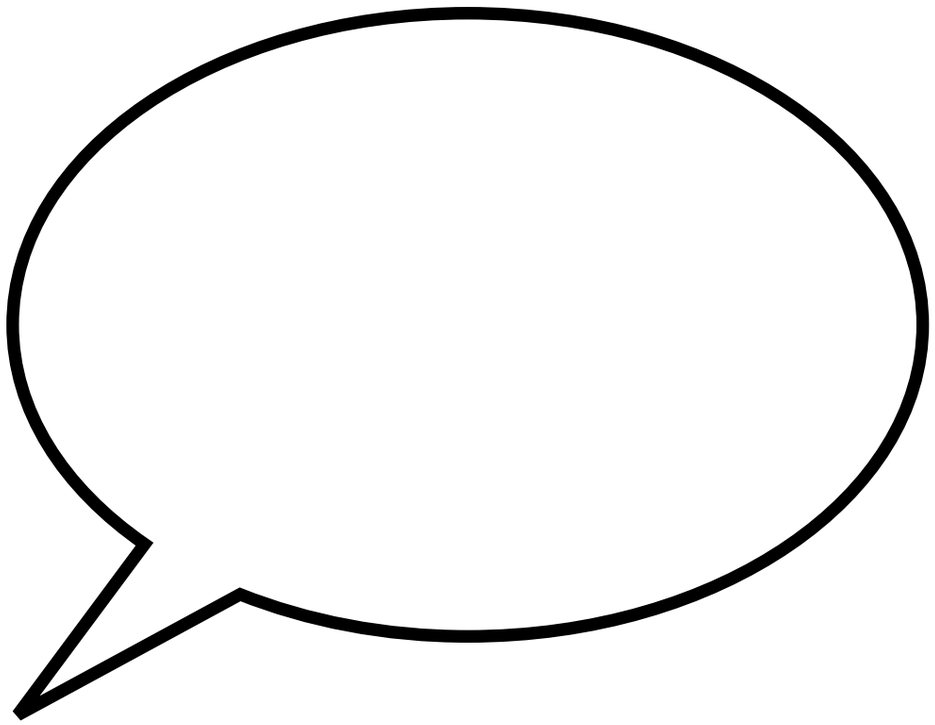
Basic Scalar Implicature Game

- The basic RSA model predicts that the pragmatic speaker will calculate the utility of each possible u as a way of conveying m.
- This is calculated as the (the natural logarithm of) the literal listener's probability of m given u, minus the "cost" of uttering u:

$$U(u,m) = \ln P_{L0}(m|u) - \text{cost}(u)$$



$P(u_{\text{all}}|m_{1-3}) = ?$



$P(u_{\text{some}}|m_{1-3}) = ?$

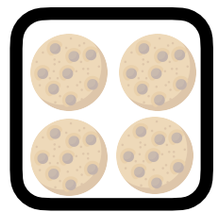
$P(u_{\text{none}}|m_{1-3}) = ?$

Basic Scalar Implicature Game

- Then this utility score is fed into the following equation to calculate $P_{S1}(u|m)$

$$P_{S1}(u|m) \propto \exp(\alpha \cdot U(u; m))$$

- Here, α is a "utility-scaling parameter" that represents how well the speaker's behavior conforms to expected utility.

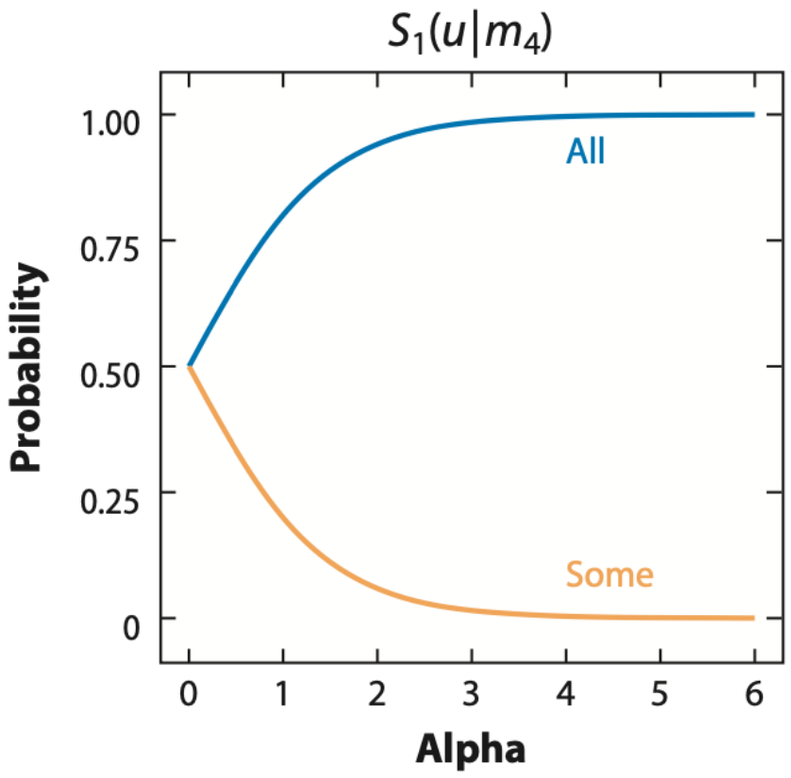


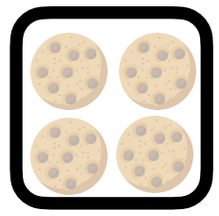
$P(u_{\text{all}}|m_{1-3}) = ?$

$P(u_{\text{some}}|m_{1-3}) = ?$

$P(u_{\text{none}}|m_{1-3}) = ?$

Alex ate some/all of the cookies.





$$P(u_{\text{all}}|m_{1\text{–}3}) = ?$$

$$P(u_{\text{some}}|m_{1\text{–}3}) = ?$$

$$P(u_{\text{none}}|m_{1\text{–}3}) = ?$$

Figure 2

Pragmatic speaker probability of using u_{some} or u_{all} to refer to m_4 under varying α, derived from the literal listener in **Figure 1*b***.

Basic Scalar Implicature Game

- Finally, consider the pragmatic listener (L1), who reasons about what the pragmatic speaker would do and updates accordingly.
- They calculate $P_{L1}(m|u)$ from $P_{S1}(u|m)$ and P(m), using Bayes' rule:

$$P_{L1}(m|u) \propto P_{S1}(u|m) \cdot P(m)$$

- Given most values for α, this winds up lowering the odds of m_4, which is the implicature we were looking for.

Alex ate some of the cookies.

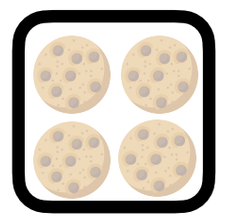
$P_{L0}(m_0|u_{some}) = 0$

$P_{L0}(m_1|u_{some}) = .25$

$P_{L0}(m_2|u_{some}) = .25$

$P_{L0}(m_3|u_{some}) = .25$

$P_{L0}(m_4|u_{some}) = .25$



Alanna

Initially, I was resistant to the idea presented by "The Rational Speech Act Framework" that there is no gray area in communication when looked at under the Baye's hypothesis. There are so many miscommunications interpersonally, between species, and with the emergence of AI. [...]

Steve

I'm curious about how we should understand interpretation as a cognitive process for RSA frameworks. Even in idealized people, do agents consider each possible meaning before inferring the most statistically likely one? This seems to require that these ideal listeners apprehend every possible meaning that could be attributed to the utterance. That seems implausible as a model of how even idealized people interpret utterances. It sounds like more of a description of deliberation and judgment than one of interpretation.

Eleonora

...Degen mentions a few difficulties for the RSA (which is a formal theory that models pragmatics according to the Bayesian conditionalization/update method), and I was surprised to not see conditional statements included in that group. I wonder whether a Bayesian approach to formal pragmatics may inherit, and thus need to account for, the notorious objections leveled against theories which make use of Bayesian reasoning in the context of conditionals (especially indicatives, wherein a conditional if A, then C corresponds to the probability of C, updated on A). In the literature on conditionals, such Bayesian reasoning is usually associated with Stalnaker's Thesis: Where A > B stands for the indicative conditional with antecedent A and consequent B, and P stands for any rational credence function such that P (A) > 0: (1)

Stalnaker's Thesis: P(A > B) = P (B | A).

Stalnaker's Thesis has been the target of quite a lot of controversy, which began when Lewis (1976) notoriously gave the first triviality results, which have since been strengthened by Fitelson (2013, 2015, 2016). Although the literature seems to have found ways to retain some version of the ST (cfr. Schulteis (2022), Bacon (2015), Goldstein and Santorio (2021), Khoo and Manderlkern (2019)), all of these proposals had to postulate some changes to the original Bayesian approach. I wonder if this would be needed for RSA too.

Cornelia

There's an objection to the Gricean story of implicature computation (see Travis, “On what is strictly speaking true”) and I think the Rational Speech Act framework (and possibly Harris's account?) might inherit it. The issue is that Gricean interpretation takes literal meaning as an input and only on that basis retrieves utterance meaning. In other words, semantic meaning is taken as basic and pragmatic meaning is derivative. But it's unclear that sentences uttered always have literal semantic meaning independent of the utterance, for on the very same circumstances, the very same sentence can be judged true or false, depending on relevance considerations. When I see you frowning over your black coffee and tell you that there's milk in the fridge, it won't be sufficient for my utterance to be true if there's a puddle of milk at the bottom. Yet when you just cleaned the fridge and I tell you there's milk in the fridge, the puddle makes my utterance true.

Now, of course, the QUD is different in the two cases. But does that help? Do we want to hold that not only pragmatic factors such as relevance but in fact the truth value of an utterance depends on the QUD? And if we do, is this sort of semantic contextualism a problem for the RSA and Harris?

Shin, responding to Cornelia

The following is my take on Cornelia's point and further thought on RSA. For the first question, I think there is no problem for contextualists to incorporate QUD since bringing QUD is a natural way to evaluate the relevance of an utterance. Indeed some recent proposals appeal to QUD in order to explain the data like in Travis' paper (Schoubye & Stokke (2016) "What is said?", Bowker (2022) "Ineliminable underdetermination and context-shifting arguments"). Also I'm not sure how this kind of approach poses any problem for RSA. Once we get the meaning of a sentence by using contextualism with QUD, this meaning will be used as input for calculating further implicated content in RSA.

NOÛS

NOÛS 00:0 (2015) 1–35
doi: 10.1111/nous.12133

What is Said?

ANDERS J. SCHOUBYE
University of Edinburgh

ANDREAS STOKKE
Umeå University

It is sometimes argued that certain sentences of natural language fail to express truth conditional contents. Standard examples include e.g. *Tipper is ready* and *Steel is strong enough*. In this paper, we provide a novel analysis of truth conditional meaning (*what is said*) using the notion of a question under discussion. This account (*i*) explains why these types of sentences are not, in fact, semantically underdetermined (yet seem truth conditionally incomplete), (*ii*) provides a principled analysis of the process by which natural language sentences (in general) can come to have enriched meanings in context, and (*iii*) shows why various alternative views, e.g. so-called Radical Contextualism, Moderate Contextualism, and Semantic Minimalism, are partially right in their respective analyses of the problem, but also all ultimately wrong. Our analysis achieves this result using a standard truth conditional and compositional semantics and without making any assumptions about enriched logical forms, i.e. logical forms containing phonologically null expressions.

Cornelia
(Degen 2022 acknowledges something like this problem when considering lexical uncertainty: in that case, "the literal listener performs the computation of literal meaning under the assumption of different possible lexicons". But lexical ambiguity is not the only way that semantic meaning can be underdetermined without context.)

Shin

Furthermore, I think RSA itself would be useful in a variety kind of cases where the meaning of an expression is semantically underdetermined. For instance, by taking the value of L as “tall” and postulating possible standards of height, the lexical uncertainty model would be expected to assign the highest probability to the value a speaker intended using “tall” (really?). The point would be that RSA can give an explanation of local pragmatic processes. And this is another departure from Grician pragmatics (though not emphasized in the paper) since Grician pragmatics basically operates on the level of sentences, not subsentential expressions. So, It seems to me that RSA can be used in the calculation of both what is said and implicature. In this sense, RSA would be a good generalization of Grician pragmatics (however, I am wondering if the pragmatic processes to derive what is said and implicature are the same).

Synthese (2017) 194:3801–3836
DOI 10.1007/s11229-015-0786-1

S.I. : VAGUENESS AND PROBABILITY

Adjectival vagueness in a Bayesian model of interpretation

Daniel Lassiter[1] · **Noah D. Goodman**[1]

Received: 31 July 2014 / Accepted: 26 May 2015 / Published online: 23 June 2015

Abstract We derive a probabilistic account of the vagueness and context-sensitivity of scalar adjectives from a Bayesian approach to communication and interpretation. We describe an iterated-reasoning architecture for pragmatic interpretation and illustrate it with a simple scalar implicature example. We then show how to enrich the apparatus to handle pragmatic reasoning about the values of free variables, explore its predictions about the interpretation of scalar adjectives, and show how this model implements Edgington's (Analysis 2:193–204,1992, Keefe and Smith (eds.) Vagueness: a reader, 1997) account of the sorites paradox, with variations. The Bayesian approach has a number of explanatory virtues: in particular, it does not require any special-purpose machinery for handling vagueness, and it is integrated with a promising new approach to pragmatics and other areas of cognitive science.

Keywords Vagueness · Probability · Cognitive science · Sorites paradox

Linguistics and Philosophy (2023) 46:1075–1130
https://doi.org/10.1007/s10988-022-09379-6

ORIGINAL RESEARCH

On the optimality of vagueness: "around", "between" and the Gricean maxims

Paul Égré[1] · **Benjamin Spector**[2] · **Adèle Mortier**[3] · **Steven Verheyen**[4]

Accepted: 9 December 2022 / Published online: 15 May 2023

Abstract
Why is ordinary language vague? We argue that in contexts in which a cooperative speaker is not perfectly informed about the world, the use of vague expressions can offer an optimal tradeoff between truthfulness (Gricean Quality) and informativeness (Gricean Quantity). Focusing on expressions of approximation such as "around", which are semantically vague, we show that they allow the speaker to convey indirect probabilistic information, in a way that can give the listener a more accurate representation of the information available to the speaker than any more precise expression would (intervals of the form "between"). That is, vague sentences can be *more informative* than their precise counterparts. We give a probabilistic treatment of the interpretation of "around", and offer a model for the interpretation and use of "around"-statements within the Rational Speech Act (RSA) framework. In our account the shape of the speaker's distribution matters in ways not predicted by the Lexical Uncertainty model standardly used in the RSA framework for vague predicates. We use our approach to draw further lessons concerning the semantic flexibility of vague expressions and their irreducibility to more precise meanings.

COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 45 (2021) e12926

ISSN: 1551-6709 online
DOI: 10.1111/cogs.12926

The Division of Labor in Communication: Speakers Help Listeners Account for Asymmetries in Visual Perspective

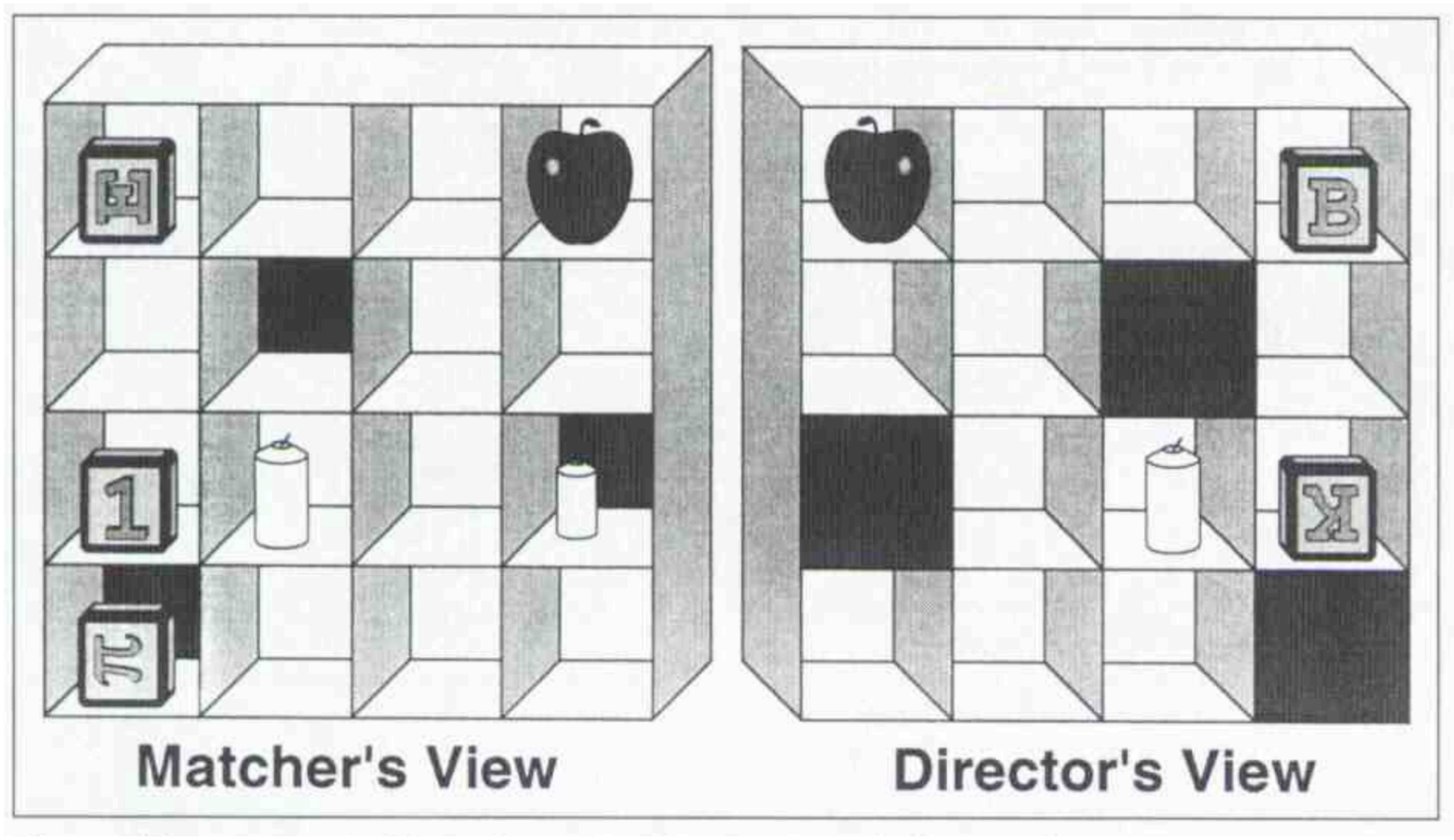
Robert D. Hawkins,[a] Hyowon Gweon,[a] Noah D. Goodman[a,b]

[a]*Department of Psychology, Stanford University*
[b]*Department of Computer Science, Stanford University*

Received 28 August 2019; received in revised form 17 September 2020; accepted 4 November 2020

THE DIRECTOR TASK

Keysar, Barr, and Horton (1998): "The Egocentric Basis of Language Use: Insights From a Processing Approach,"



Director's instructions to Matcher:
"Put **the bottom block** below the apple."

If the Matcher moves the block marked ⊏, then they have reasoned "egocentrically"—i.e., failed to account for the Director's perspective.

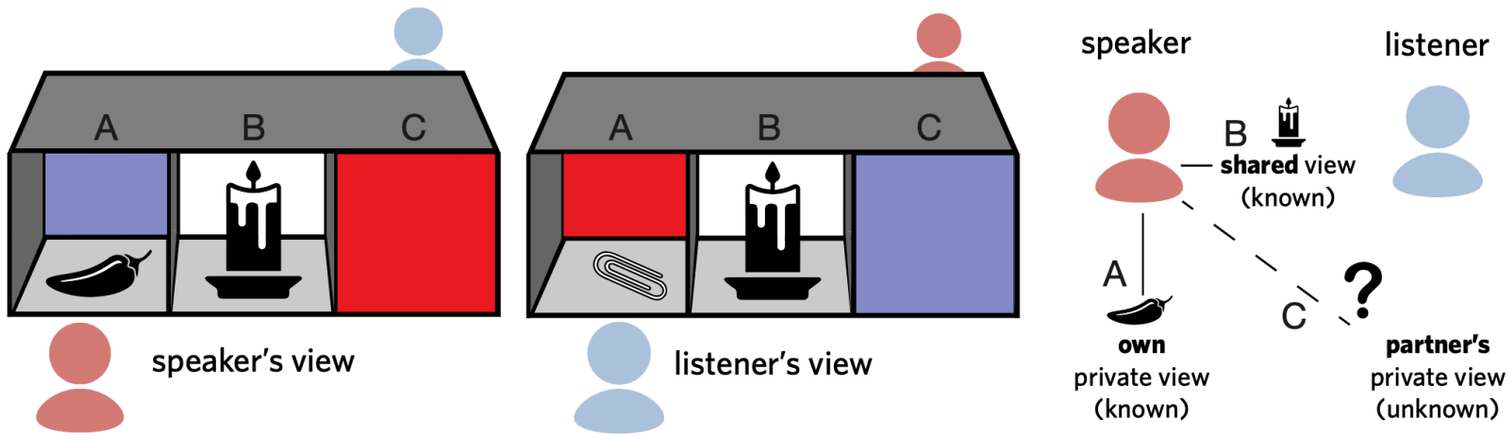
PATTERNS OF BREAKDOWN

Speakers and hearers are often sensitive to others' perspectives.

But not always. Some patterns:

- cognitive load → more egocentric (Keysar 2008)
- Verbal-working-memory deficit → more egocentric (Lin et al 2010)
- Time constraints → more egocentric (Horton and Keysar 1996)
- Younger children → more egocentric (Keysar 2008)
- Repeated conversations with egocentric interlocutor → less egocentric (Hawkins et al 2008)

- Speakers compensate for uncertainty about addressees' perspectives by using more informative descriptions (Hawkins et al 2021)



Theo (on the Hawkins reading)

Something I'm curious about though is how to interpret the cost function. In the model we only have cost of an utterance, but in real conversation, especially if we think of conversations extending to many participants and across a long time, the 'opportunity cost' of being unclear, e.g. communicating with insufficient exactness the spot you wanted a couch moved to to one person leading to the need to later call another and incur all cognitive costs of communication again, seems at least as important. I wonder if this is a partial reason for participants in experiment 1 behaving so close the max perspective-taking effort; that it's a habit to over-optimize for clarity on each utterance due to downstream effects? This multi-utterance cost could further exacerbate the issues relating to the cost of estimating cost brought up in the discussion as well.

Hawkins et al on Resource Rationality

The recent development of resource- rational analysis (Griffiths, Lieder, & Goodman, 2015; Lieder & Griffiths, 2019; Shenhav et al., 2017) has provided a framework for understanding a range of costly but important cognitive functions, including attention (Padmala & Pessoa, 2011), working memory maintenance (Howes, Duggan, Kalidindi, Tseng, & Lewis, 2016), planning (Callaway et al., 2018), and decision-making under uncertainty (Lieder, Griffiths, & Hsu, 2018), through the application of rational principles under cognitive constraints. Computational- level accounts are often under-constrained: There are many solutions to the computational problem that could be considered equally "optimal" a priori regardless of how costly or intractable the required computations are. Resource-rational analyses attempt to place stronger constraints on these accounts by incorporating processing considerations. The key insight, motivated by recent work on the mechanisms of cognitive control, is that agents consider both the functional value of a computation as well as its costs (Kool & Botvinick, 2018; Shenhav, Botvinick, & Cohen, 2013), and behave in a way that is consistent with an approximately optimal trade-off between them.

Hawkins et al on Resource Rationality

We consider the trade-off between one specific benefit (the expected value of communicative accuracy) and one specific cost (the cognitive cost of perspective-taking). (p.11)

If communicative accuracy were the only consideration, it would always be preferable to use maximal perspective-taking..., since higher perspective-taking leads to higher accuracy. In a resource-rational model, however, these benefits are traded off against the costs of perspective-taking. For simplicity, we assume that cost is linear in the degree of perspective-taking.... [F]or now, we maintain an abstract notion of "cost" encompassing multiple processing considerations.... (p.11)

Hawkins et al on Resource Rationality

Our theoretical framework relies on an abstract computational notion of "effort" or "cost." We remain agnostic about the precise source of these costs at the algorithmic level; the director–matcher task, like many other standard tasks used to evaluate theory of mind abilities (Quesque & Rossetti, 2020), involves the coordination of many cognitive systems, and the available data do not allow us to isolate a specific cause for poor perfor- mance (Rubio-Ferna´ndez, 2017). We expect that the abstract cost associated with using a higher mixture weight in our model represents a range of different costs associated with general executive control, working memory, selective attention, and other processes, as well as whatever cost may be specifically associated with forming and maintaining repre- sentation of a partner's likely behavior given their perspective. (p.32)

Sadie (on the Hawkins reading)
I'd like to talk more about how we might understand a cognitive or perceptual perspective, and what in particular the idea of 'perspective-taking' brings to the table. Thinking about perspectives I'm always inclined to go to Elisabeth Camp's work, like what she says here in 'Perspectives in Imaginative Engagement with Fiction':

'Trying on a perspective involves more than imagining an experience or the truth of a set of propositions: it requires actually structuring one's intuitive thinking in the relevant way [...] so that certain sorts of properties stick out as especially notable and explanatorily central in one's intuitive thinking'.

Harris talks in chapter 2 about using information about the addressee's 'perceptual and cognitive perspective' in communication design, but I hadn't thought so much up until reading the Hawkins about this kind of perspective-sensitivity as involving a process of 'trying on' like Camp describes. Although we don't engage imaginatively in ordinary conversation in the same way, this characterization of our engagement with fictional perspectives makes me curious about how exactly we do utilise information about the cognitive perspectives of others to communicate effectively. I am reminded of how I might adapt my descriptions of the same item during a game of Articulate, depending on which member of my family I am playing with – although I don't imaginatively engage with their differing cognitive perspectives exactly, different sorts of properties do intuitively stick out as explanatorily central...

Cognition 210 (2021) 104618

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Mindreading in conversation

Evan Westra [a,*], Jennifer Nagel [b]

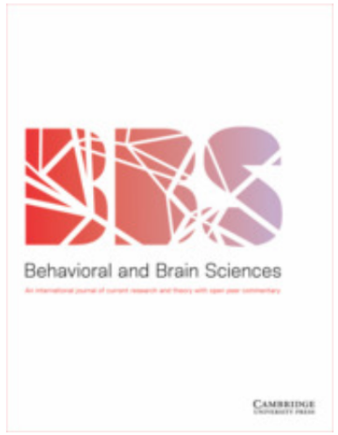
[a] Department of Philosophy, York University, Canada
[b] Department of Philosophy, University of Toronto, Canada

ARTICLE INFO

Keywords:
Mentalizing
Conversation
Factivity
Knowledge
Decoupling

ABSTRACT

How is human social intelligence engaged in the course of ordinary conversation? Standard models of conversation hold that language production and comprehension are guided by constant, rapid inferences about what other agents have in mind. However, the idea that mindreading is a pervasive feature of conversation is challenged by a large body of evidence suggesting that mental state attribution is slow and taxing, at least when it deals with propositional attitudes such as beliefs. Belief attributions involve contents that are decoupled from our own primary representation of reality; handling these contents has come to be seen as the signature of full-blown human mindreading. However, mindreading in cooperative communication does not necessarily demand decoupling. We argue for a theoretical and empirical turn towards "factive" forms of mentalizing here. In factive mentalizing, we monitor what others do or do not know, without generating decoupled representations. We propose a model of the representational, cognitive, and interactive components of factive mentalizing, a model that aims to explain efficient real-time monitoring of epistemic states in conversation. After laying out this account, we articulate a more limited set of conversational functions for nonfactive forms of mentalizing, including contexts of meta-linguistic repair, deception, and argumentation. We conclude with suggestions for further research into the roles played by factive versus nonfactive forms of mentalizing in conversation.



Behavioral and Brain Sciences

Knowledge before belief

Published online by Cambridge University Press: **08 September 2020**

Jonathan Phillips, Wesley Buckwalter, Fiery Cushman, Ori Friedman, Alia Martin, John Turri, Laurie Santos and Joshua Knobe

Show author details

Abstract

Research on the capacity to understand others' minds has tended to focus on representations of *beliefs,* which are widely taken to be among the most central and basic theory of mind representations. Representations of *knowledge*, by contrast, have received comparatively little attention and have often been understood as depending on prior representations of belief. After all, how could one represent someone as knowing something if one does not even represent them as believing it? Drawing on a wide range of methods across cognitive science, we ask whether belief or knowledge is the more basic kind of representation. The evidence indicates that nonhuman primates attribute knowledge but not belief, that knowledge representations arise earlier in human development than belief representations, that the capacity to represent knowledge may remain intact in patient populations even when belief representation is disrupted, that knowledge (but not belief) attributions are likely automatic, and that explicit knowledge attributions are made more quickly than equivalent belief attributions. Critically, the theory of mind representations uncovered by these various methods exhibits a set of signature features clearly indicative of knowledge: they are not modality-specific, they are factive, they are not just true belief, and they allow for representations of egocentric ignorance. We argue that these signature features elucidate the primary function of knowledge representation: facilitating learning from others about the external world. This suggests a new way of understanding theory of mind – one that is focused on understanding others' minds in relation to the actual world, rather than independent from it.

Keywords

DOI: 10.1111/mila.12157

WILEY

ORIGINAL ARTICLE

Factive and nonfactive mental state attribution

Jennifer Nagel

Department of Philosophy, University of Toronto, Toronto, ON, Canada

Correspondence
Department of Philosophy, University of Toronto, 170 St George Street, Toronto, Canada.
Email: jennifer.nagel@utoronto.ca

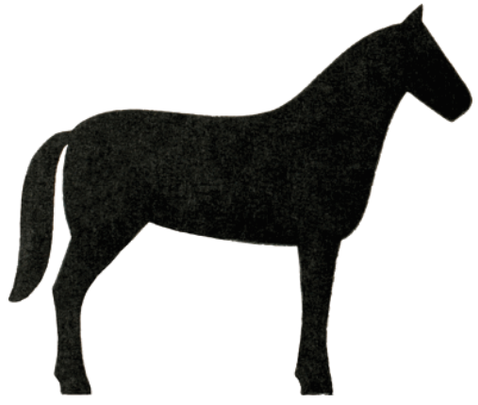
Factive mental states, such as knowing or being aware, can only link an agent to the truth; by contrast, nonfactive states, such as believing or thinking, can link an agent to either truths or falsehoods. Researchers of mental state attribution often draw a sharp line between the capacity to attribute accurate states of mind and the capacity to attribute inaccurate or "reality-incongruent" states of mind, such as false belief. This article argues that the contrast that really matters for mental state attribution does not divide accurate from inaccurate states, but factive from nonfactive ones.

KEYWORDS

belief, factivity, knowledge, mental state attribution, mindreading

Factive vs. Non-Factive Mindreading

How did *he* get in here?



Stalnaker (1978)

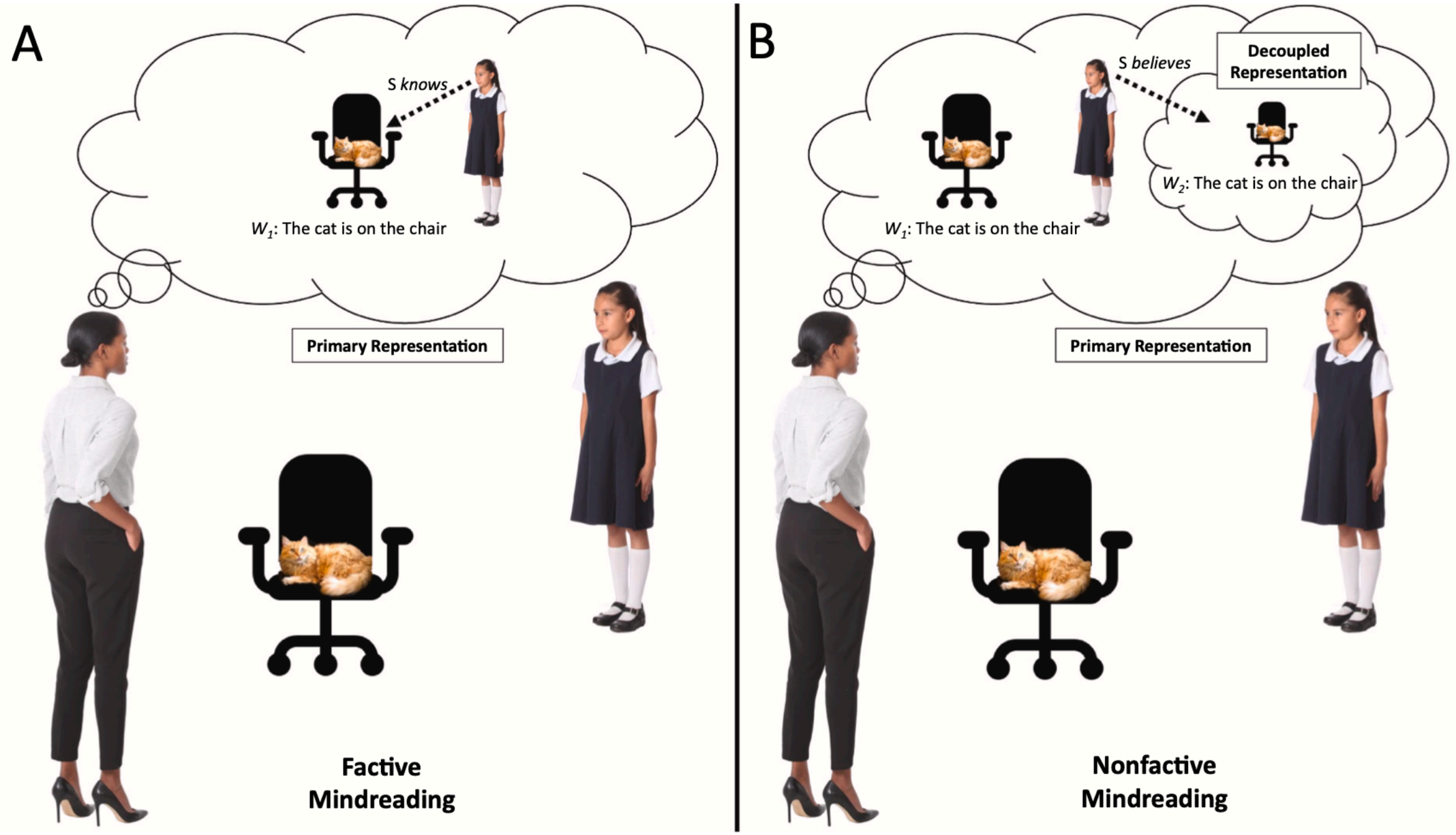
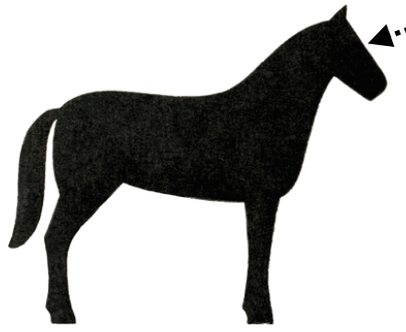


Fig. 1. Factive and nonfactive mindreading in a basic case of knowledge attribution. 1A: In factive mindreading, the content of the mental state imputed to S is *coupled* with the content of the attributor's primary representation W_1. 1B: In nonfactive mindreading, the content imputed to S is *decoupled* from the attributor's primary representation W_1, and is instead represented as a distinct representational token W_2.

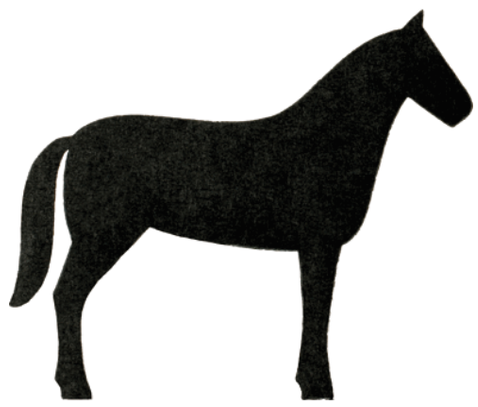
Westra and Nagel (2022)

my colleague knows

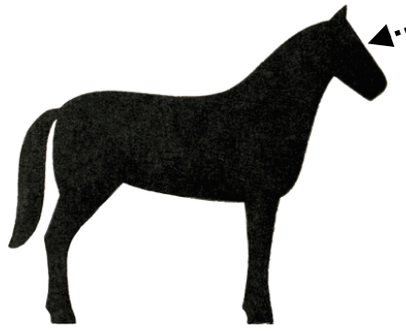


w1: The horse is in the room.

How did *he* get in here?

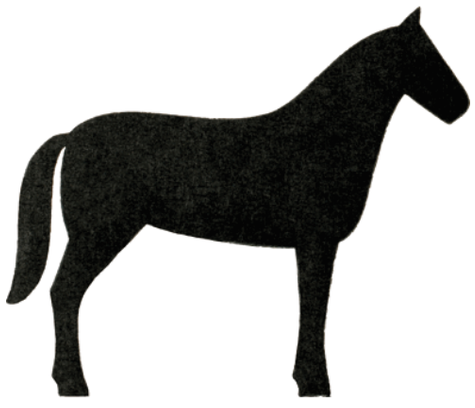


my colleague doesn't know

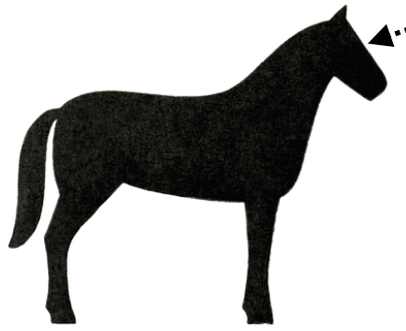


w1: The horse is in the room.

Look, a horse!

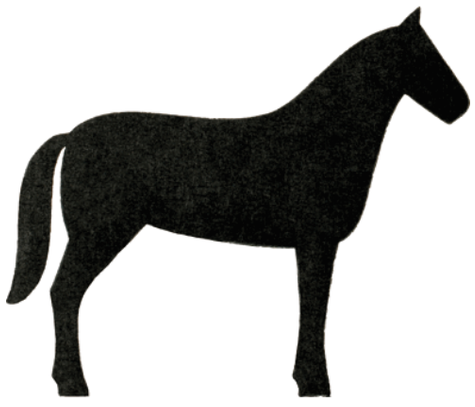


my colleague knows what the horse's name is



w1: The horse is named Secretariat.

How did Secretariat get in here?



Elliot

To summarize briefly for those who read other stuff, Suppose A and B are both looking at a cat. A represents the cat as part of her model of reality. According to the authors, there are two distinct ways A may attribute a representation of the cat to B. In factive mindreading, A will couple B's model of reality to her own and represent that B knows there's a cat. In non-factive mindreading A decouples B's model of reality from her own and attributes a belief to B that may or mat not align with A's own model of reality. The claim is that factive mindreading is (relatively) effortless and frequently used in conversation whereas non-factive mind reading is effortful and used more rarely.

Elliot

(1) What exactly does it mean to 'couple' someone else's representations to your own model of reality? It doesn't seem like A can simply attribute to B the same model of reality that A has. In the cat example, presumably A and B will see the cat from visual different perspectives so A can't attribute precisely same model of reality to B. But then what exactly does A coupled from her primary representation to B? Won't determining what should be coupled involve the same sort of reasoning as non-factive mindreading?

Griffin, responding to Elliot:

W&N have some comments that might answer your question (1). They suggest that in factive mindreading, S attributes to A the knowledge of S's primary representation p, including S's mode of presentation/ level-two features of p (p. 6). However, I think there is another way W&N can go here: instead of holding that S's factive mindreading attributes to A knowledge of S's primary representations including their mode of presentation/level-2 features, perhaps S attributes to A knowledge of S's primary representations excluding their mode of presentation/level-2 features. This seems potentially more plausible. But then we need to say more about how one can access one's own primary representations at a level that abstracts from their modes of presentation/level-two features.

Steve, responding to Elliot:
A swing at (1): intentions are beliefs about what one will do. These beliefs are the conclusions of practical reasoning, and in some sense, when you intend to phi, you must represent yourself as having knowledge about what you'll do (or try to do).

Griffin, responding to Steve (slight paraphrase):
That doesn't sound like what Bratman would say! 🙀

Steve, responding to Griffin (slight paraphrase):
Not everyone agrees with Bratman bro. 🙄

Elliot

(3) This model seems related to the Spinozan Model of belief fixation (we initially accept everything we think and rejecting is a later effortful process). I'd be interested to see how the developmental time tables line up for rejecting our own beliefs, and performing nonfactive mindreading.

Griffin (Elliot also had a similar Q)

W&N frame their proposal as a way of making intentionalists' posited unconscious processes more cognitively realistic. In doing so, they focus on the factive mental state of knowledge. However, intentionalists view the central mindreading process in conversation to be intention recognition. Intentions are not clearly factive mental states. So, it's unclear how W&N's proposal makes the intentionalists' hypothesis any more palatable, given that W&N have explained a way that factive mental states attribution can be cognitively easy, while intentionalists posit (nonfactive) unconscious and fast intention attribution.

Two possible answers. (1) Intentions are a sort of factive mental state. I have no idea how this would go... (2) W&N argue that while intentionalists focus on intention attribution, they also acknowledge that "communication design" and audience interpretation involve mindreading mental states other than intentions – such as knowledge. So, perhaps W&N's proposal works for this part of intentionalists' mindreading. However, this leaves untouched the posit of intention attribution as a cognitively demanding process that intentionalists think is at work in (at least many) conversations. So the question of whether this is psychologically realistic remains.

Petru
two possible points of tension between [W&N's] picture and Harris'.

(a) If communication normally involves the use of our advanced planning capacity, then it will require at least the following three elements: (1) representations of the world as it is (primary representations); (2) representations of the world as we want it to be after our plans are carried out; (3) decoupled representations of the communicative partner's own plans and subplans – which are not connected with any present state of the world, but rather with a future state. However, Westra and Nagel's model submits that the use of such decoupled representations is much more computationally demanding, and so much less frequent than the alternative – factive mindreading. Part of the disagreement seems to be based on different approaches to the computational cost objection. Westra and Nagel acknowledge the force of the "Too costly to be real" counter made by opponents of mentalizing and, by way of response, propose a way to make the picture more psychologically real. Harris agrees that our picture should be psychologically real, but disagrees that the mentalizing picture as it stands is too costly to be real and hints at an argument to that effect (see below). I'd be curious to hear some direct replies to the authors (e.g. Andrews, 2012; Gallagher, 2001) who "[assign mentalizing] a sharply limited function" (Westra & Nagel, p. 1).

Petru

(b) Harris claims that, even though intention recognition and the use of advanced planning in communication entail high computational costs, the relative benefits that they provide to our communicative aptitude make it worth it. There is an evolutionary story undergirding this position, a hint of which we have seen in Chapter 3 – "the advantages of cooperative communication and other forms of joint action are so significant that they constitute a powerful selection pressure in favor of cooperation-conducive traits," one of which is mindreading (Harris, Ch. 3, p. 5). However, Westra and Nagel also appeal to an evolutionary explanation for our factive mindreading abilities and their widespread deployment (i.e. gaze-tracking facilitates knowledge attribution, the human eye is particularly well-designed biologically for easy gaze-tracking, so it's plausible that our visual organs adapted to make "visually based communication" easier). But the two accounts are at odds with one another – Harris would not, I think, want to grant the distinction between factive and nonfactive mindreading too readily; or if he does, he would want to assign different weights to the two types of mindreading than Westra and Nagel do – so Westra and Nagel's evolutionary claim stands in need of reply. Furthermore, I'd be curious to hear a little more about the evolutionary claim in Harris – a plausible and detailed genealogical story for our mentalizing abilities would go a long way, I think, towards arguing that the mentalizing camp is not imputing computationally implausible demands on the cognitive systems deployed in communication.

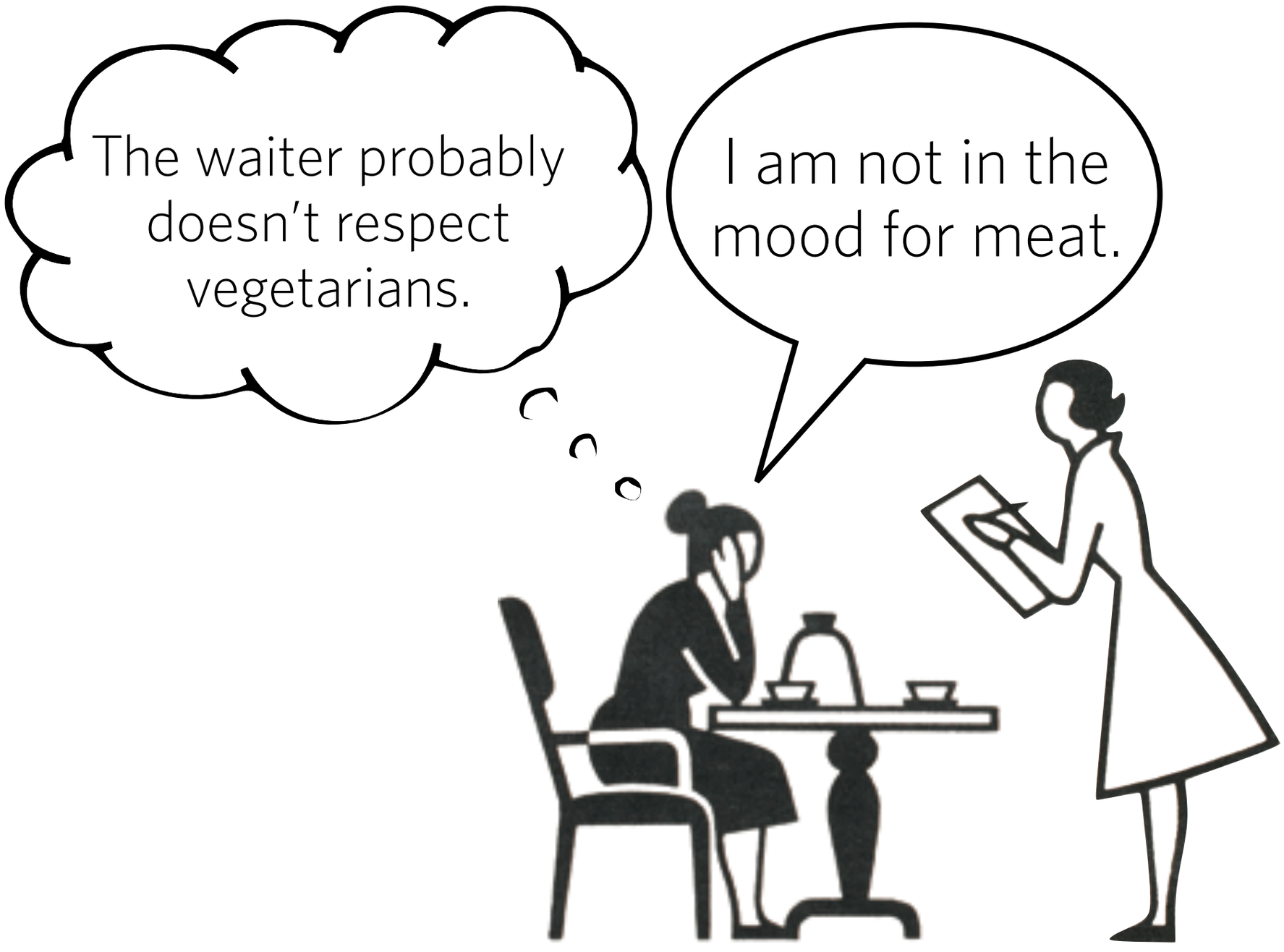
When do we need non-factive mindreading?

- Deliberate deception
- Epistemic vigilance
- Metalinguistic repair
- Argumentation

Westra and Nagel (2020)

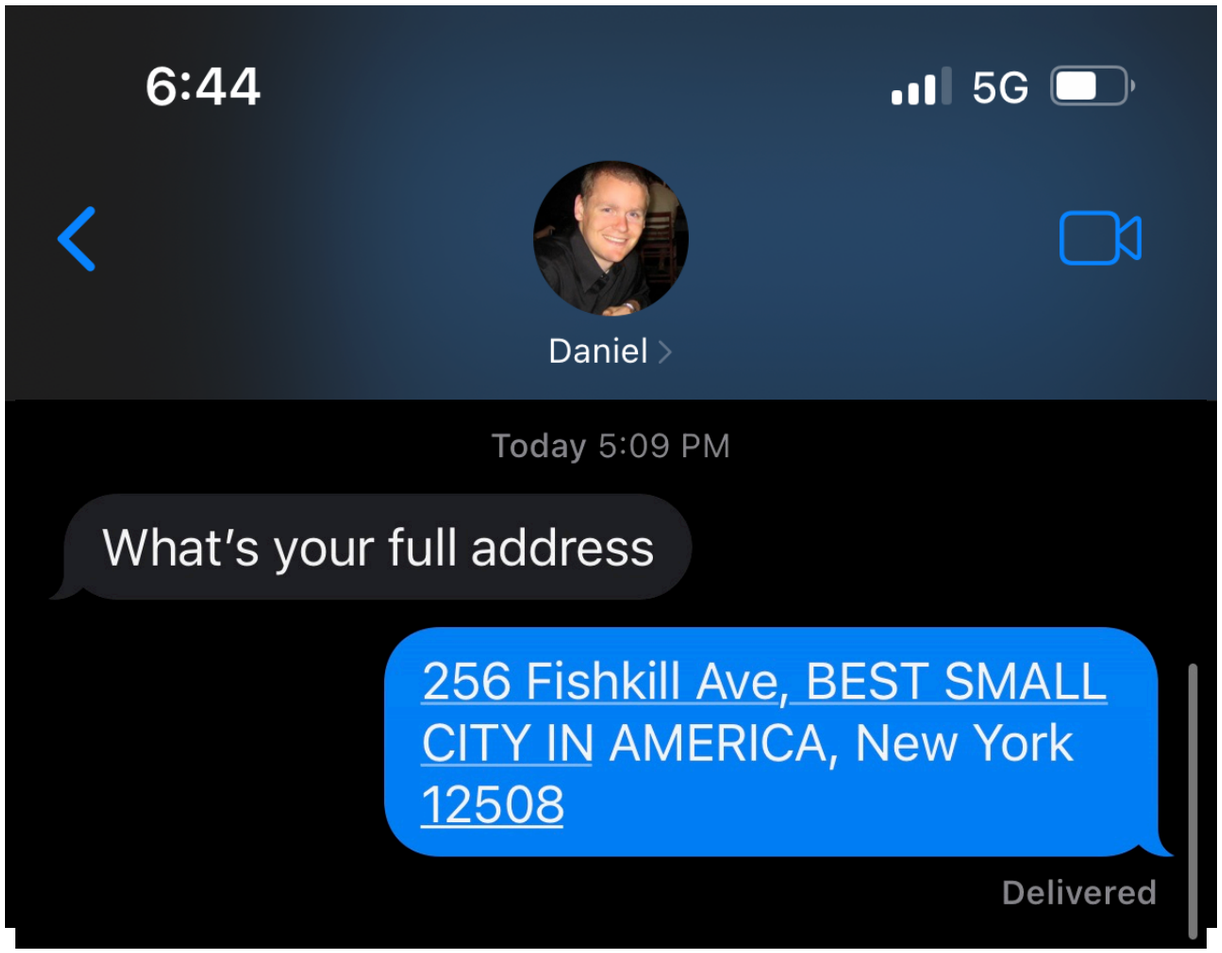
All Background Disagreement

(Not just argumentation)



All Background Disagreement

(Not just argumentation)



Oona knows that I'm not being serious,
and she'll find this fun.

This book was written by
Blippy. He's my favorite philosopher.

Pretense

Oona knows that I'm not being serious,
and she'll find this fun.

This book was written by
Blippy. He's my favorite philosopher.

Pretense

Small Talk

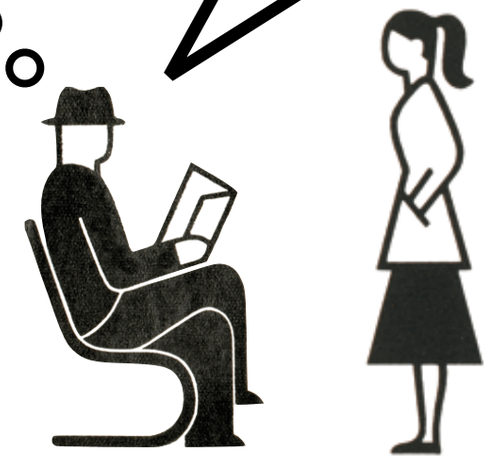
If I say, “The Investigations,” Oona won’t understand. But if I say more, she might.

This is one of my philosophy books. It’s by a philosopher named Ludwig Wittgenstein. It’s called *The Philosophical Investigations*.

Prediction

My colleague wants/intends hire Job Candidate A, but I want to hire Candidate B.

Did you see Candidate A's reference letter from Professor X?



Non-epistemic mindreading

Interpersonal Object Tracking


He is coming to the party tonight.

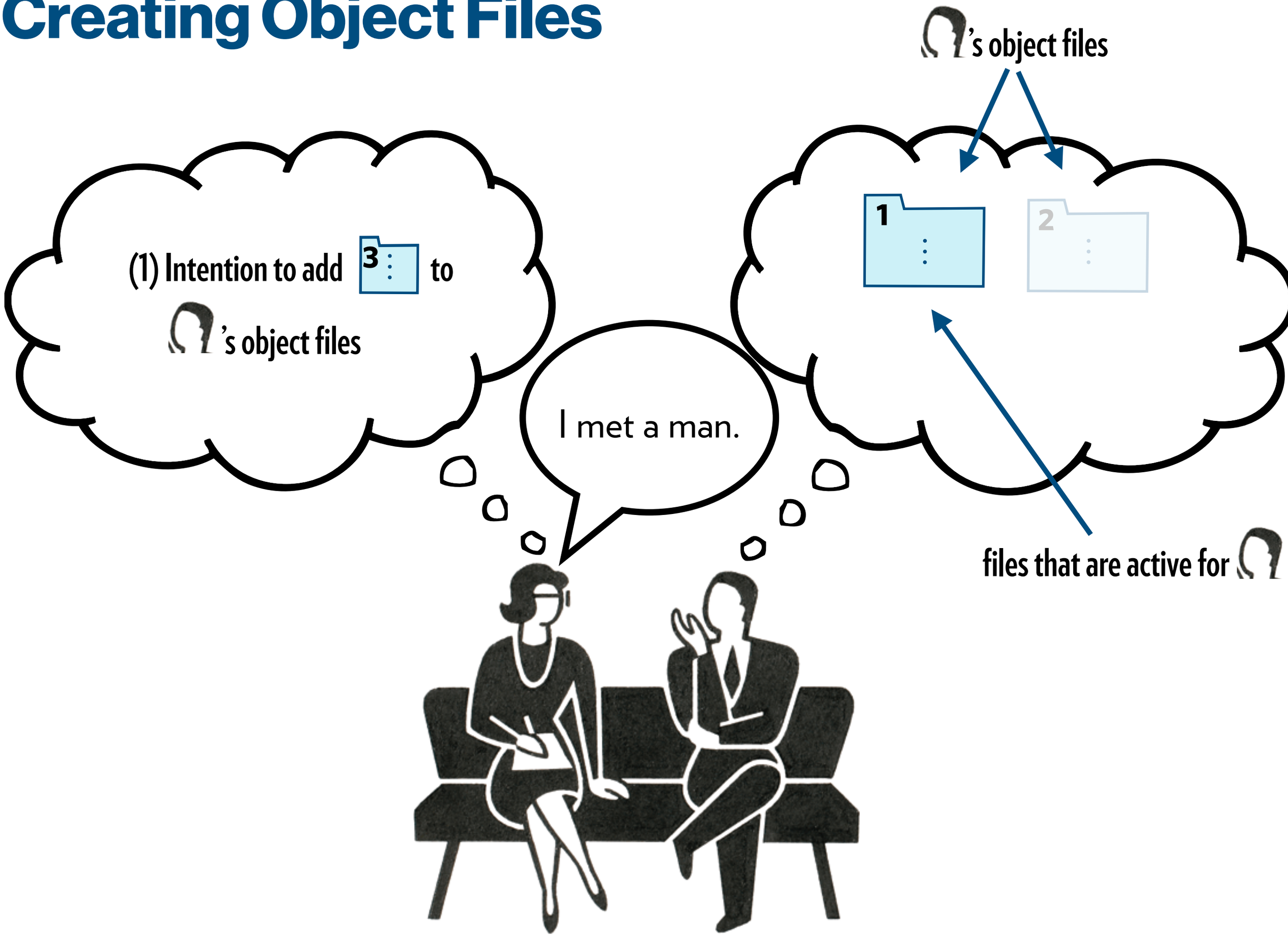
I know a man.

He is coming to the party tonight.

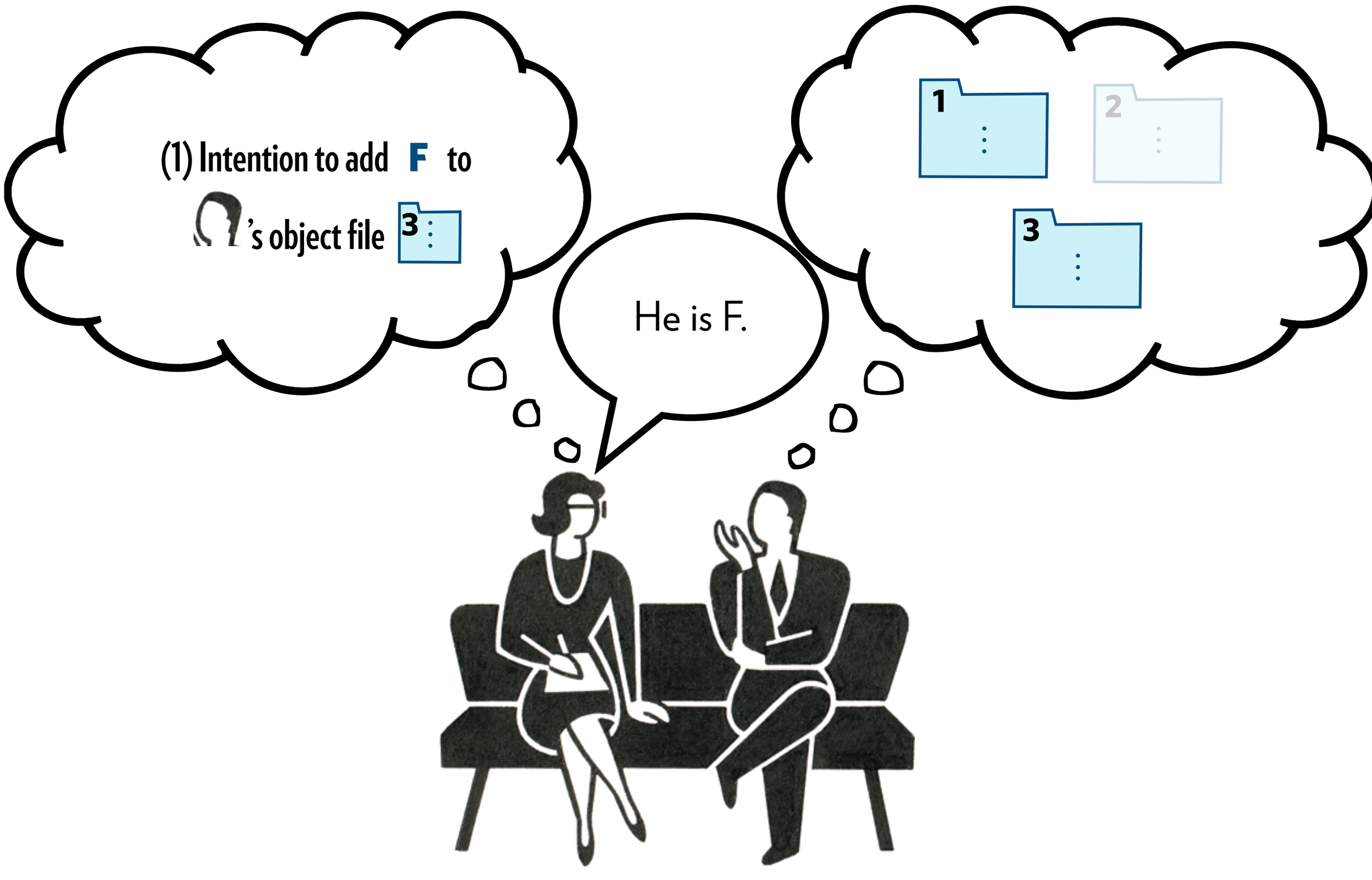
I propose that the common ground of a context be identified with what I have been calling the "file" of that context. As we will see, files cannot be construed as sets of possible worlds, although each file determines such a set.

—Heim (1982)

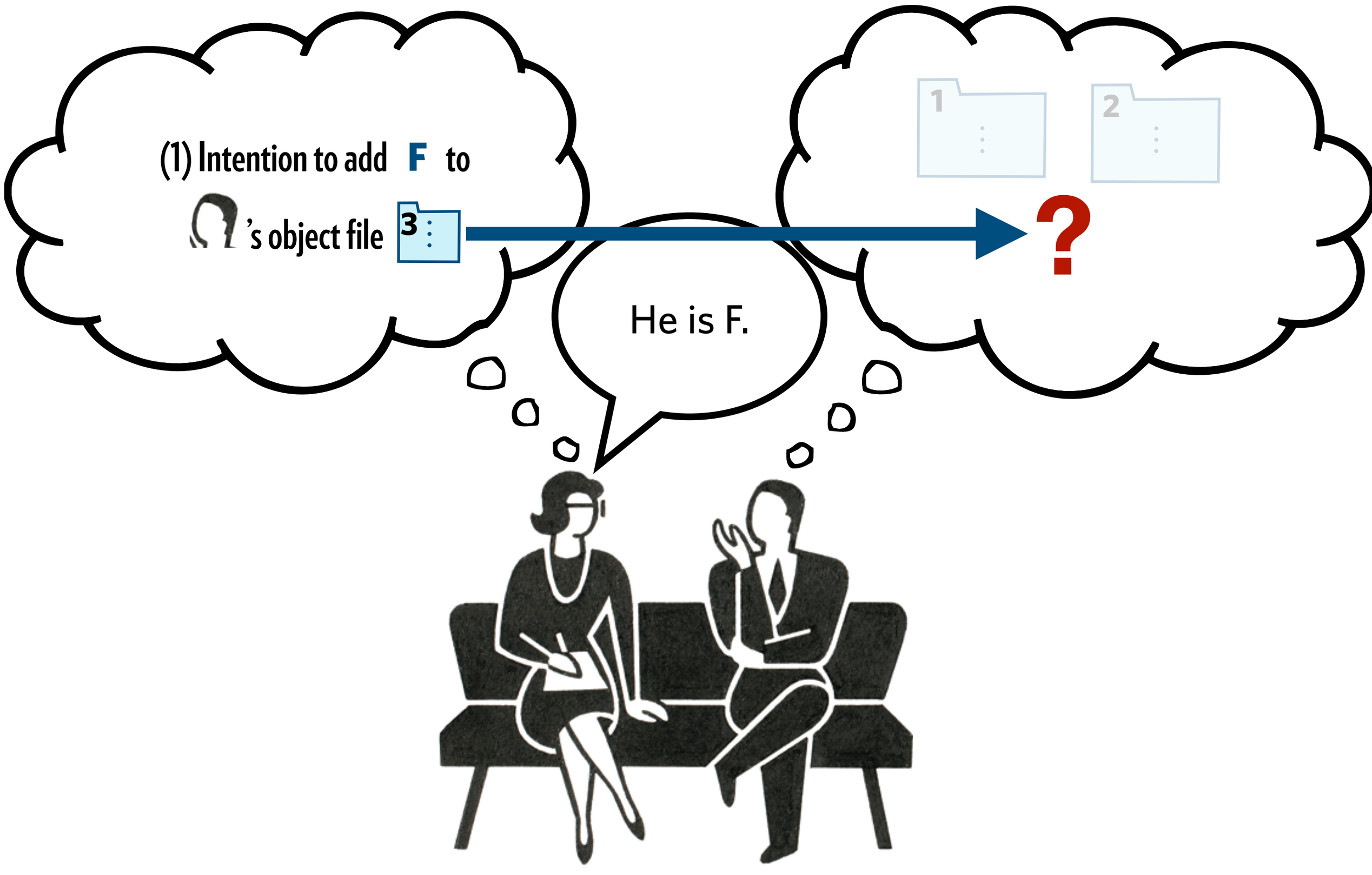
Creating Object Files



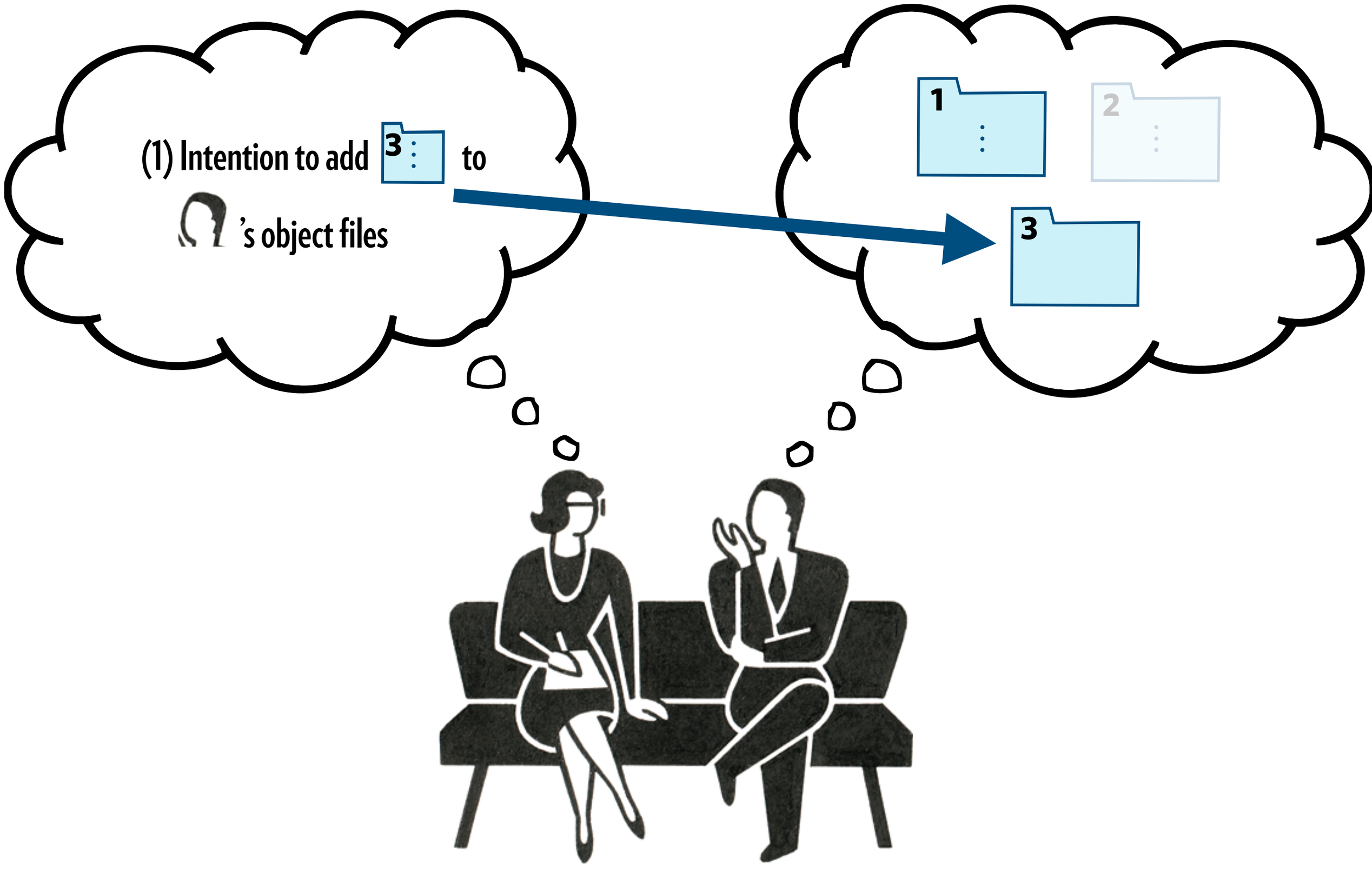
Updating Object Files



Presupposition Failure



Interpersonal File Identification?



Interpersonal File Identification?