LINGUISTIC COMMUNICATION & SOCIAL COGNITION

PETER VAN ELSWYK
NORTHWESTERN

DANIEL HARRIS
CUNY GRADUATE CENTER, HUNTER COLLEGE

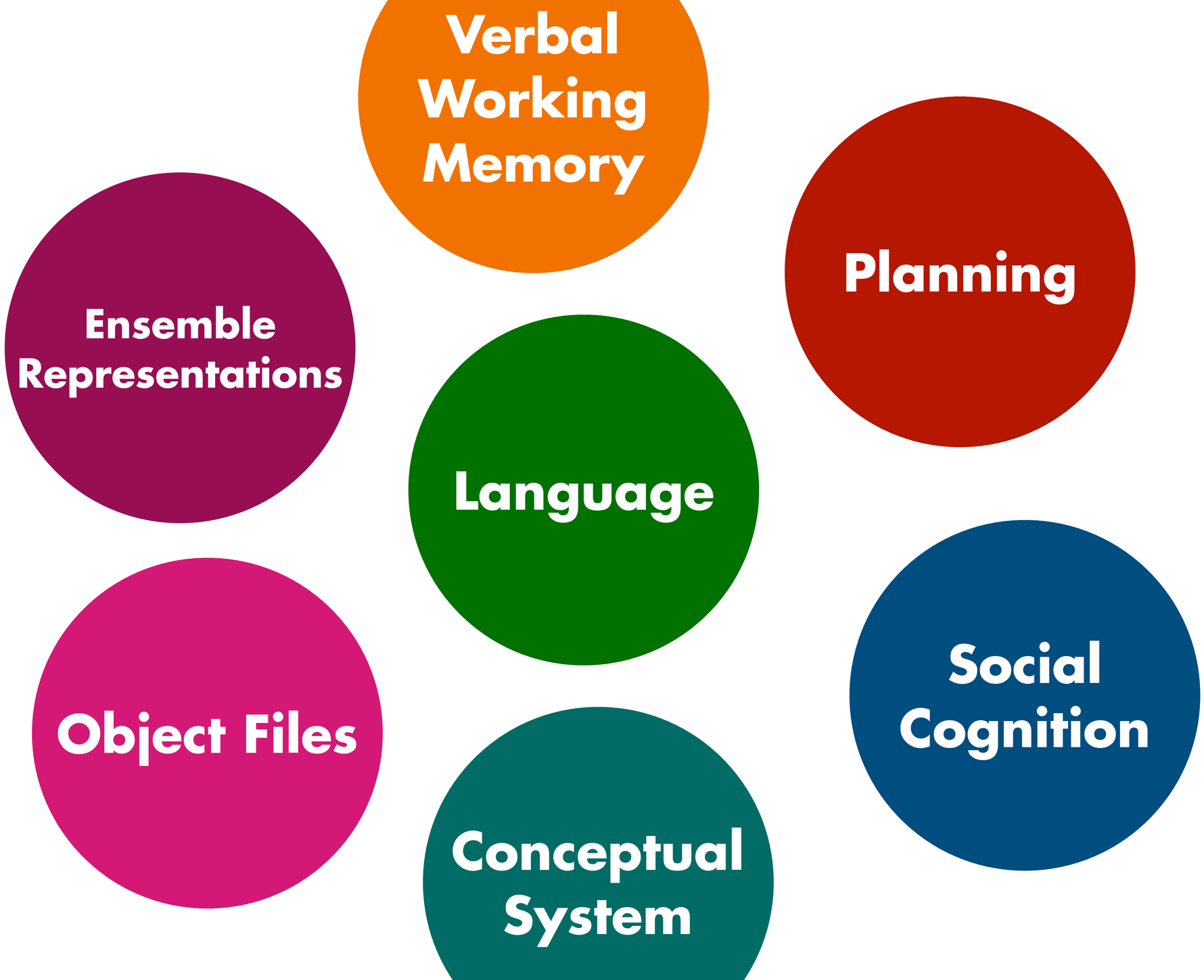
DAY 1:
INTRODUCTION

PETER VAN ELSWYK
NORTHWESTERN

DANIEL HARRIS
CUNY GRADUATE CENTER, HUNTER COLLEGE

WHY SHOULD YOU (A NASSLLI STUDENT) CARE ABOUT SOCIAL COGNITION?


Language


Verbal Working Memory
Planning
Ensemble Representations
Language
Social Cognition
Object Files
Conceptual System

Different quantificational determiners tend to make use of different representational systems.

Verbal Working Memory

Planning

Conceptual System

Social Cognition

Language

⟦each F is G⟧ = (∀x: Fx)(Gx)

⟦every F is G⟧ = [The X : F](X)[(∀x: Xx)(Gx)]

Object Files

This system is used to represent small numbers of things, considered as individuals.

Ensemble Representations

This system is used to represent summary properties of collections, taken as wholes.

Contents lists available at ScienceDirect

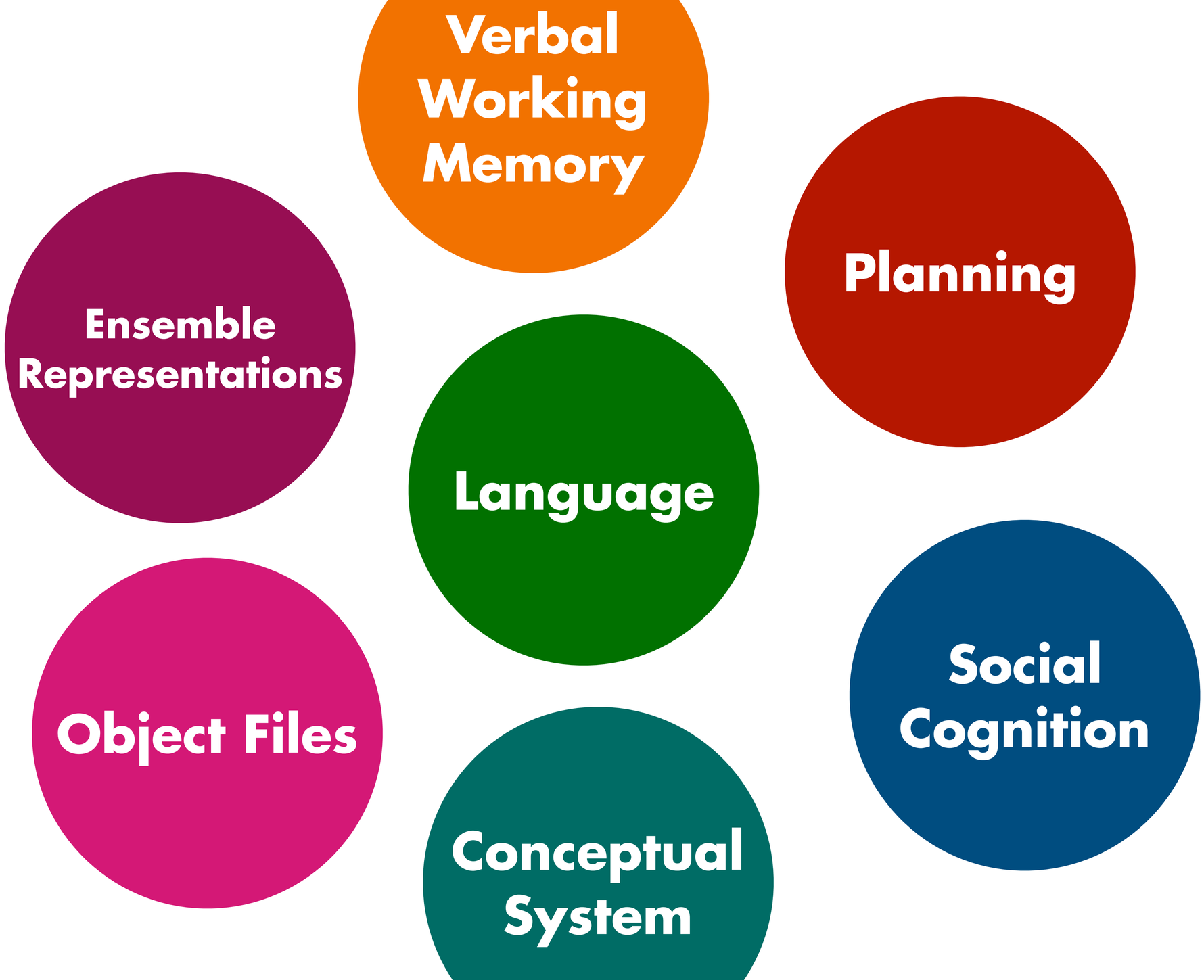
Cognitive Psychology

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

Keeping quantifier meaning in mind: Connecting semantics, cognition, and pragmatics

Tyler Knowlton [a,*], John Trueswell [b], Anna Papafragou [c]

[a] MindCORE, University of Pennsylvania, 3740 Hamilton Walk, Philadelphia, PA 19104, USA
[b] Department of Psychology, University of Pennsylvania, 425 South University Ave, Philadelphia, PA 19104, USA
[c] Department of Linguistics, University of Pennsylvania, 3401 Walnut Street, Philadelphia, PA 19104, USA


Verbal Working Memory
Planning
Ensemble Representations
Language
Social Cognition
Object Files
Conceptual System

Social Cognition

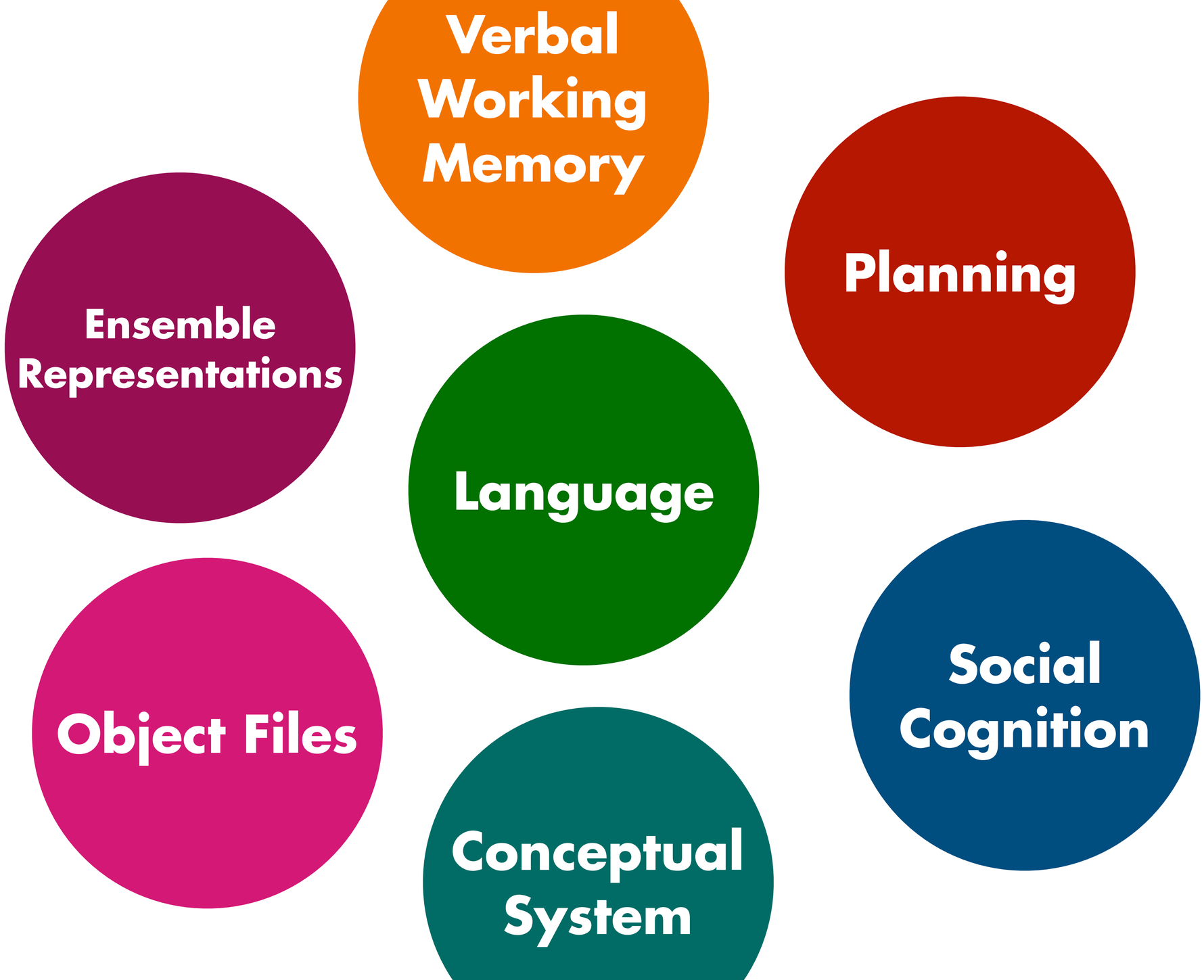
"…the various psychological processes that enable individuals to take advantage of being part of a social group."

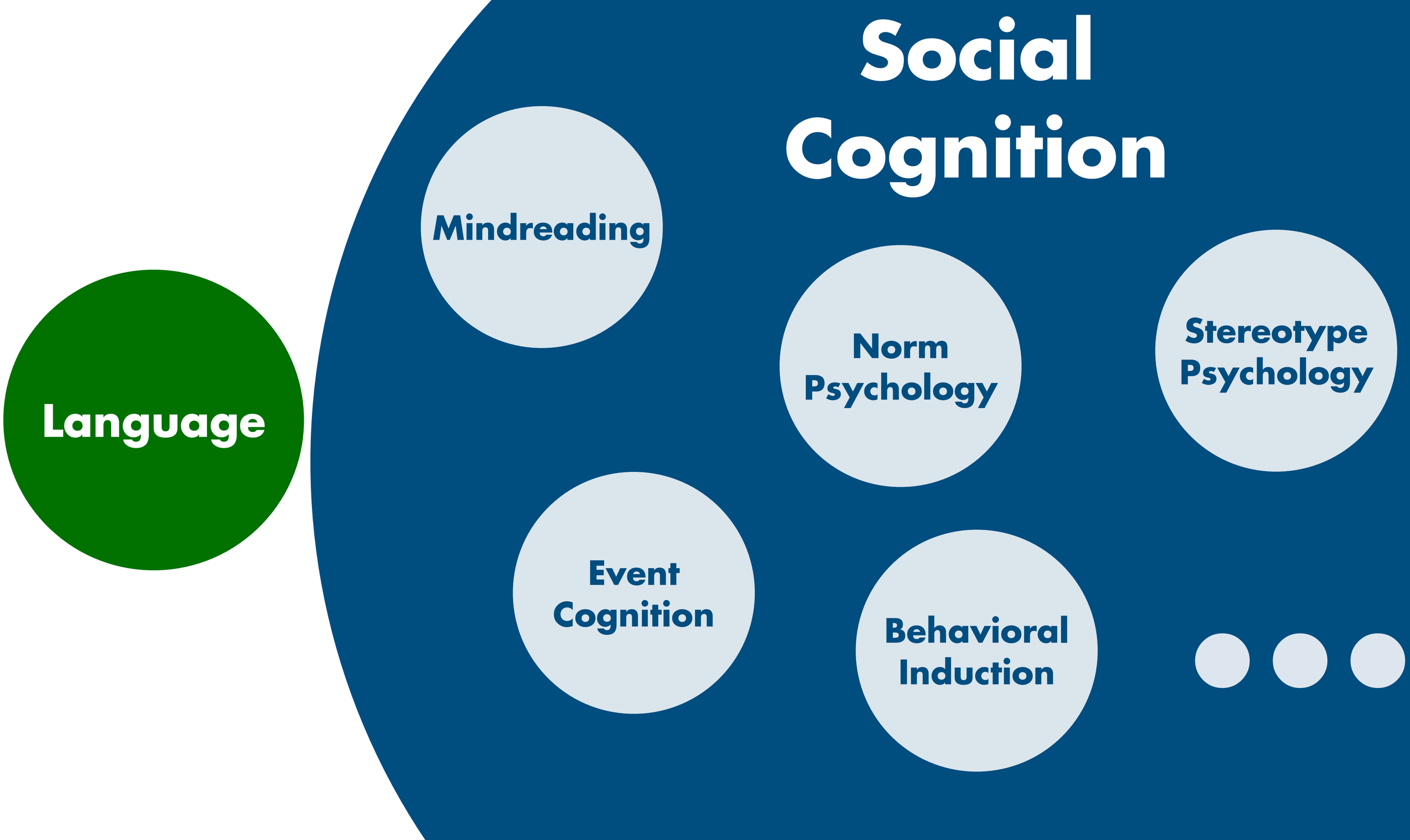
—Chris Frith

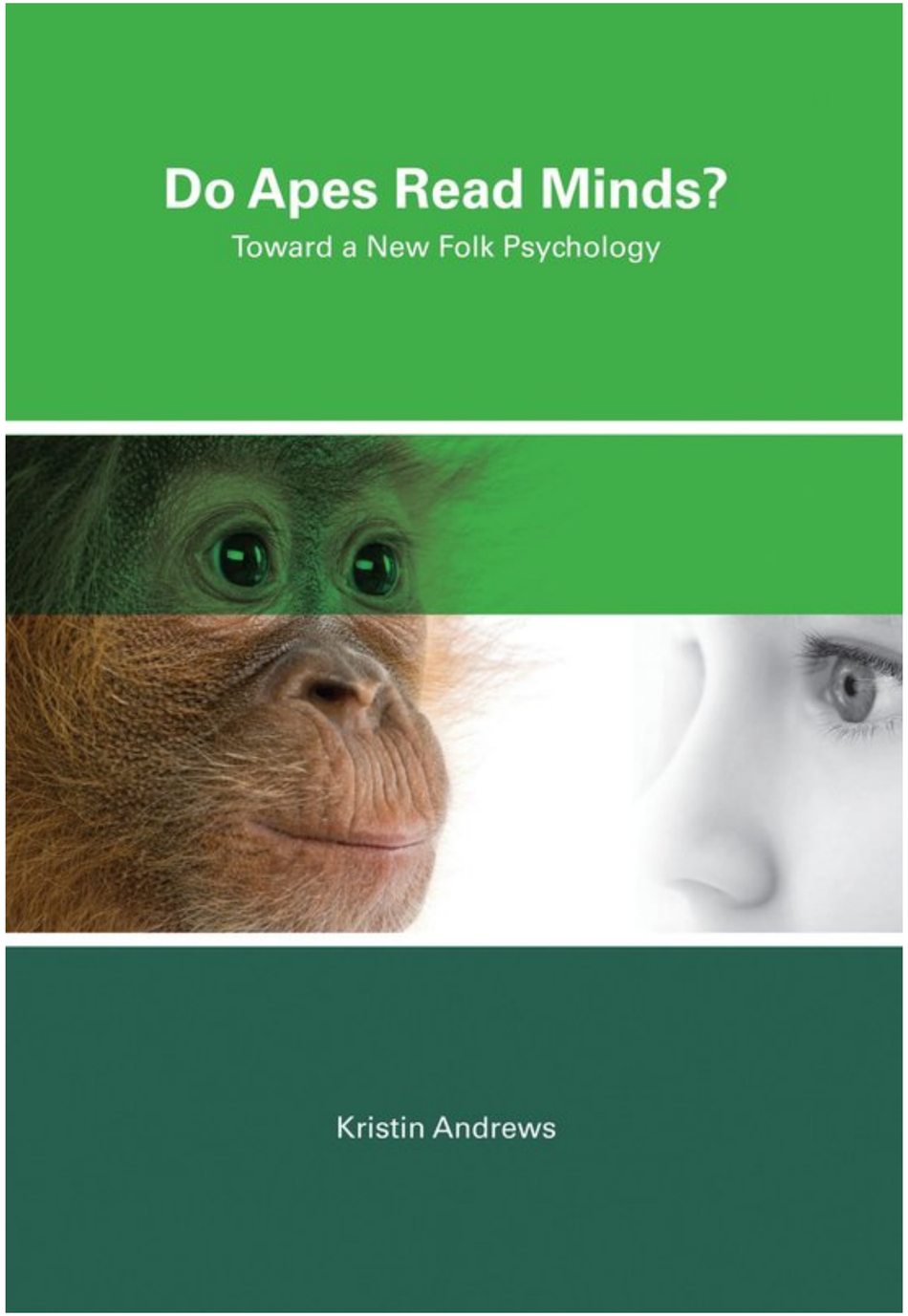
"Social Cognition," Philosophical Transactions of the Royal Society B, 2008

Social Cognition

The collection of cognitive capacities that allow us to understand and predict other agents, and to coordinate our own actions with theirs.


Verbal Working Memory
Planning
Ensemble Representations
Language
Social Cognition
Object Files
Conceptual System

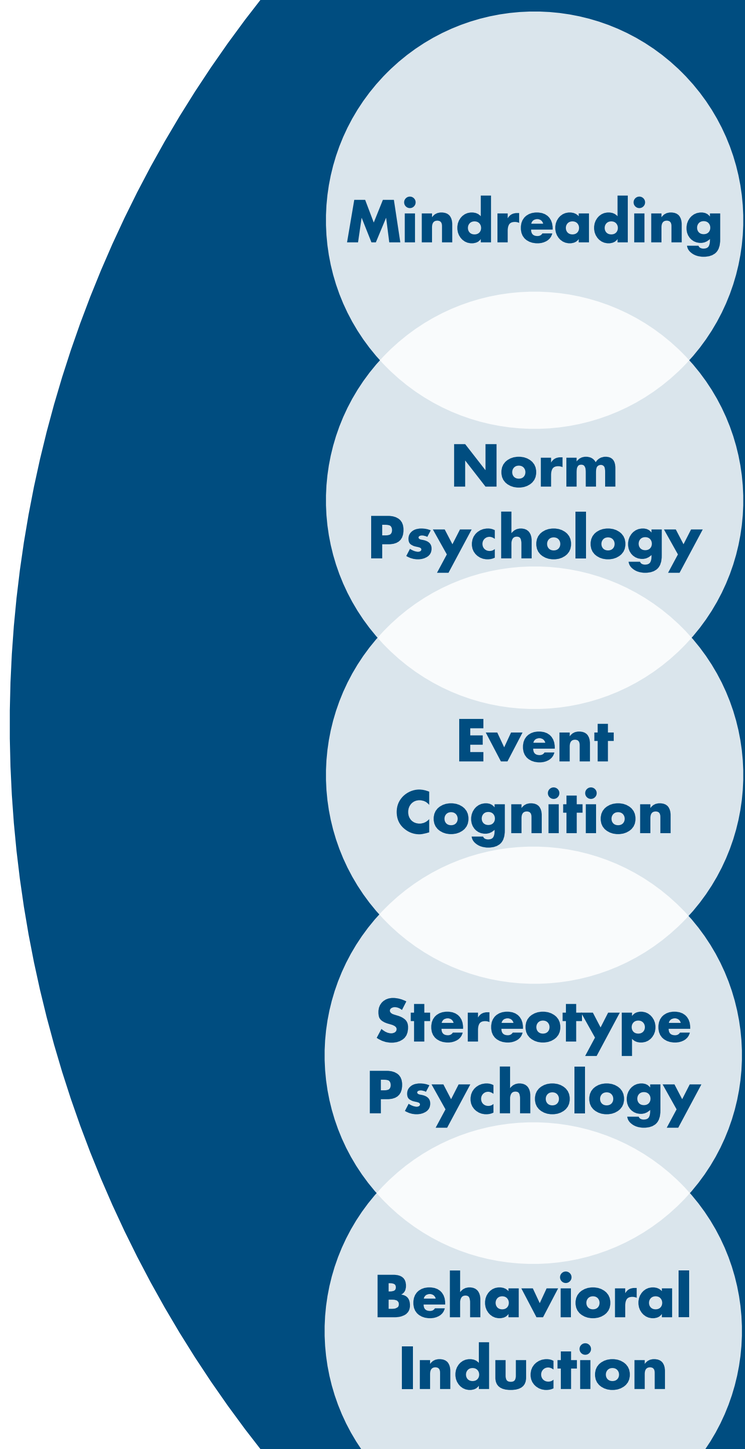

Social Cognition
Mindreading
Norm Psychology
Stereotype Psychology
Language
Event Cognition
Behavioral Induction

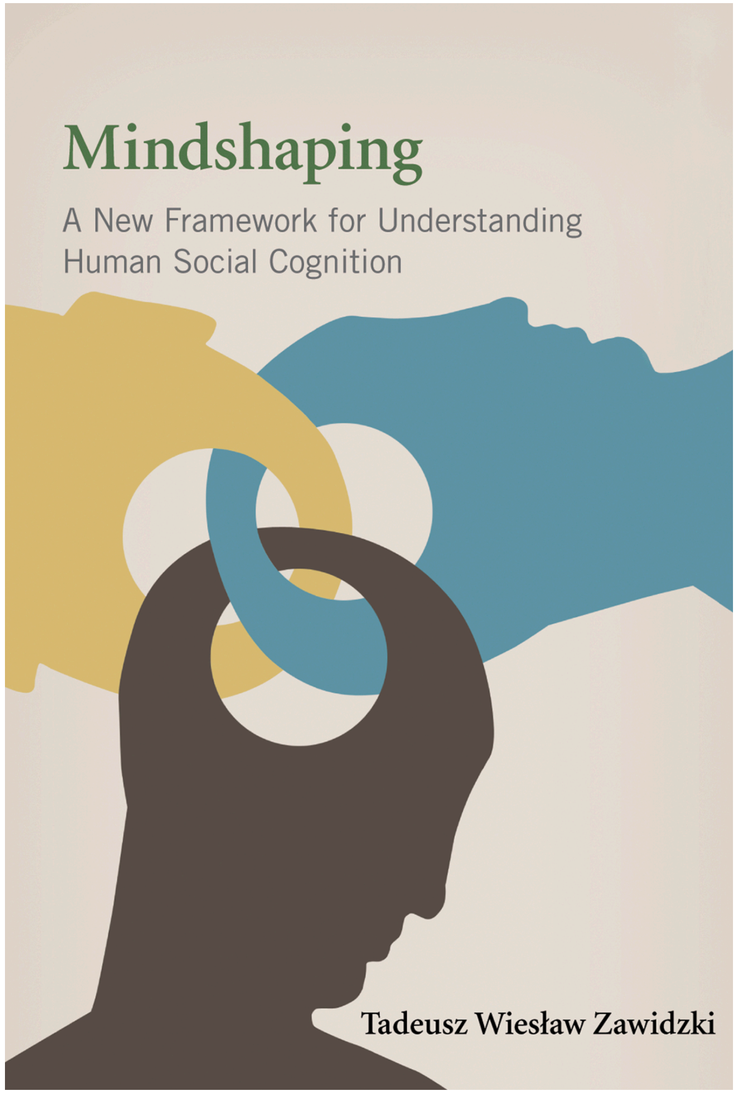


- Mindreading
- Norm Psychology
- Event Cognition
- Stereotype Psychology
- Behavioral Induction

...in our quotidian predicting, explaining, and coordinating behavior, we don't need to read minds. Rather than mindreading, we use a host of different cognitive strategies. Some of these, such as predicting that others will do what we tend to do (Krueger 1988) or relying on stereotypes and social roles to predict that people will do what they should do as members of society (Locksley et al. 1980), can be used with people we don't know at all. When we have close relationships with persons, we can use other strategies, such as inductive generalizations over one's past behavior (Kalish 2002), primary intersubjectivity (Trevarthen 1979), or trait attributions (Nisbett & Ross 1991). We also do attribute propositional attitudes, of course, but only in addition to other non-propositional mental states such as emotions and intentions.

—Kristin Andrews
Precis of *Do Apes Read Minds?* (2013)





PHILOSOPHICAL PERSPECTIVES

Philosophical Perspectives, 32, Philosophy of Language, 2018
doi: 10.1111/phpe.12114

PROVINCIALISM IN PRAGMATICS

Josh Armstrong
UCLA

0. Introduction

What is the explanatory scope of pragmatic theory outside the context of uses of language in general, and of linguistic communication in particular?

This question has typically been approached by considering the extent to which the concepts and models of pragmatics are needed to get a theoretical grip on non-linguistic devices of human social interaction—for instance, in acts of ostension or pantomiming, in the use of pictures and diagrams, in acts of dance or the like. In what follows, I will approach the question from a distinct but complementary angle. My focus will be on the extent to which pragmatics is needed to get a theoretical grip on the non-linguistic devices of *non-human* social interactions—for instance, the vocalizations of monkeys, the rituals of pair-bonded birds, or the gestures of apes.

The central claim of my paper is that pragmatics has a wider scope of application than has been generally appreciated. In particular, I will argue that many discussions of pragmatics are guilty of a problematic form of *provincialism*. The provincialism at issue restricts the class of target systems of study to

Mindreading

Norm Psychology

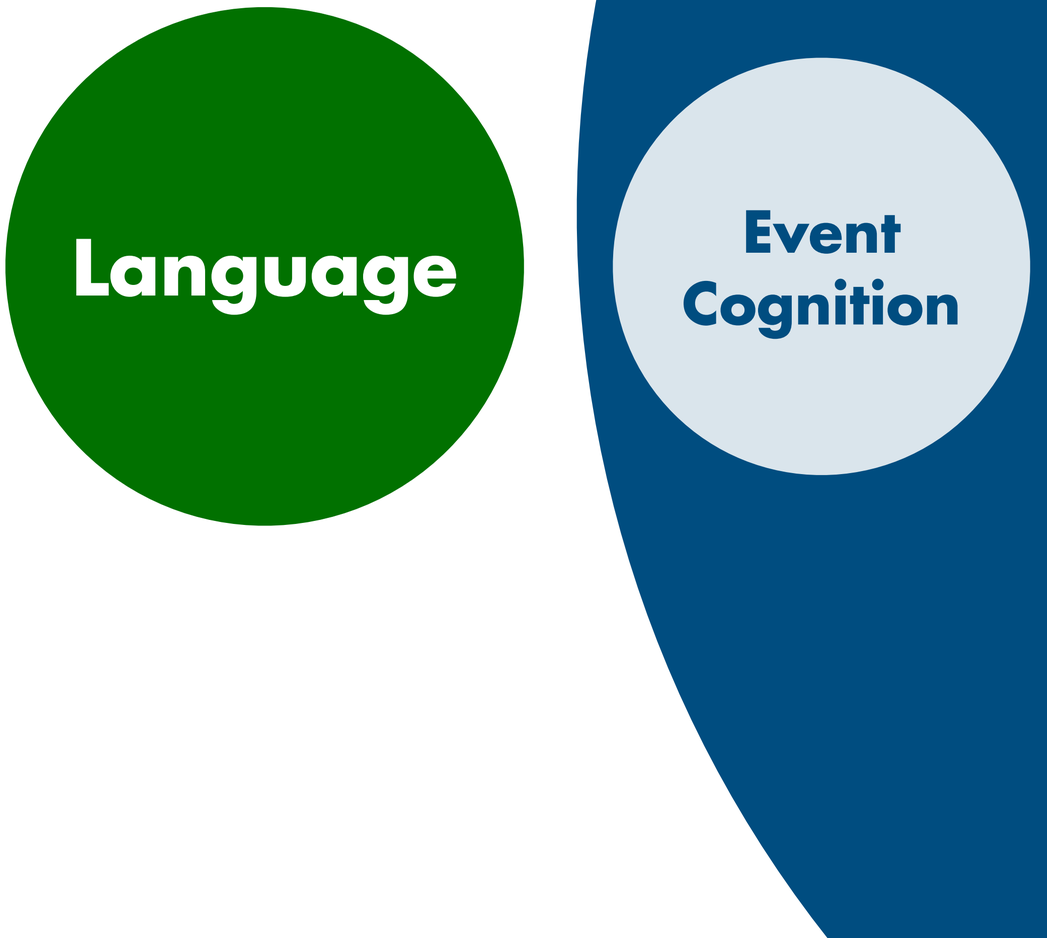
Event Cognition

Stereotype Psychology

Behavioral Induction

- A domain-general capacity for representing experienced or hypothetical *events*.
- Event representations are constructed by identifying their boundaries, i.e., the end of one event is the beginning of another.
- Events are individuated by *location*, *participants*, *kind*, and *goal*. Changes in these properties typically constitute boundaries.

Event Cognition


Language
Event
Cognition


Language
“Shane organized NASSLLI.”
“Shane was organizing NASSLLI.”


Event
Cognition

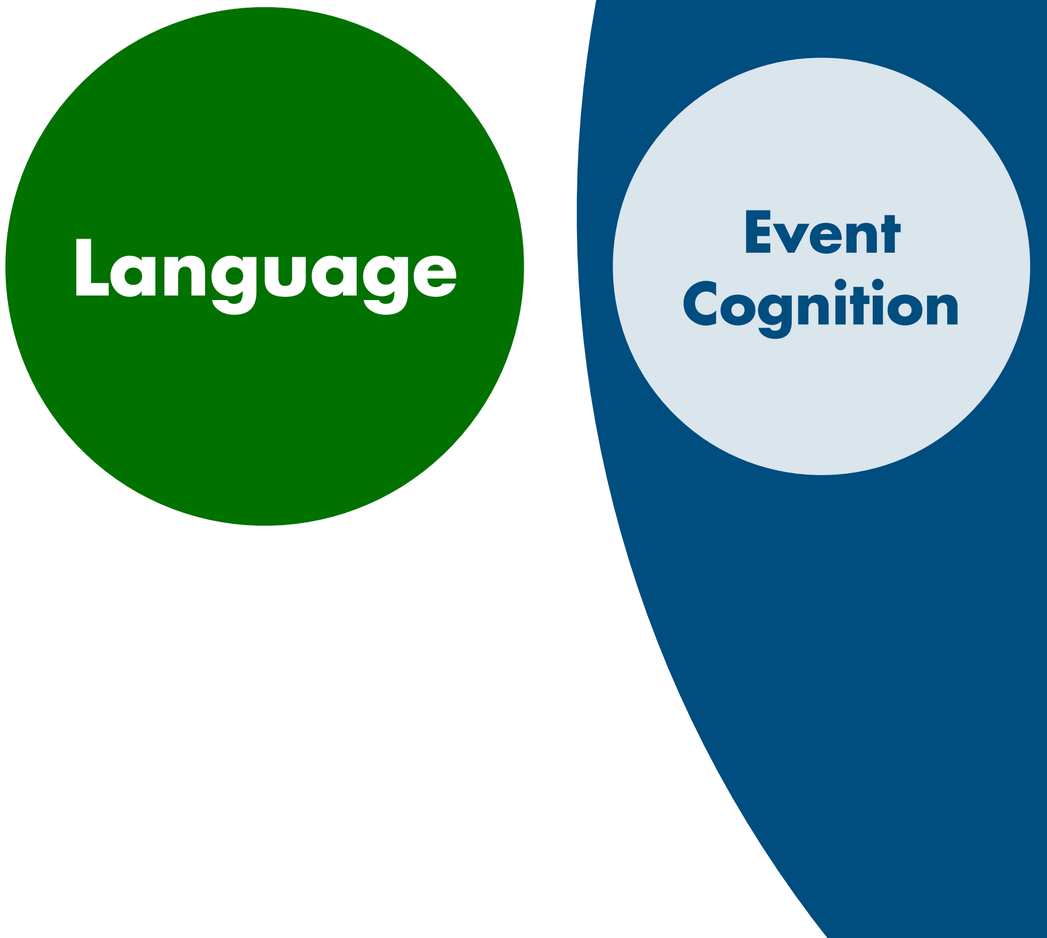
Language

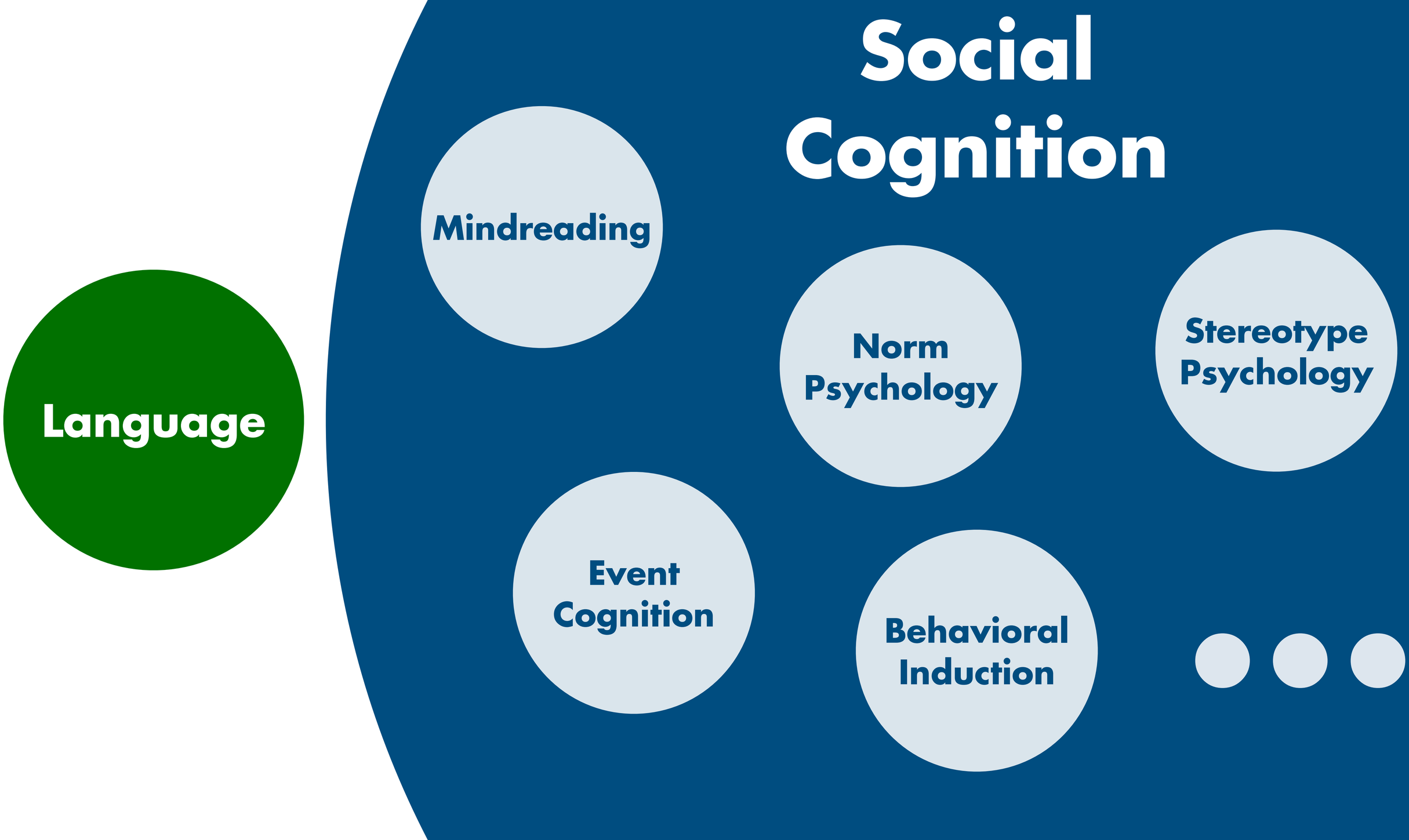
"Shane organized NASSLLI."

"Shane was organizing NASSLLI."

MUCH MORE ON WEDNESDAY!

Event Cognition


Language
Event
Cognition


Social Cognition
Mindreading
Norm Psychology
Stereotype Psychology
Event Cognition
Behavioral Induction
Language

Mindreading

Language

Belief Attribution

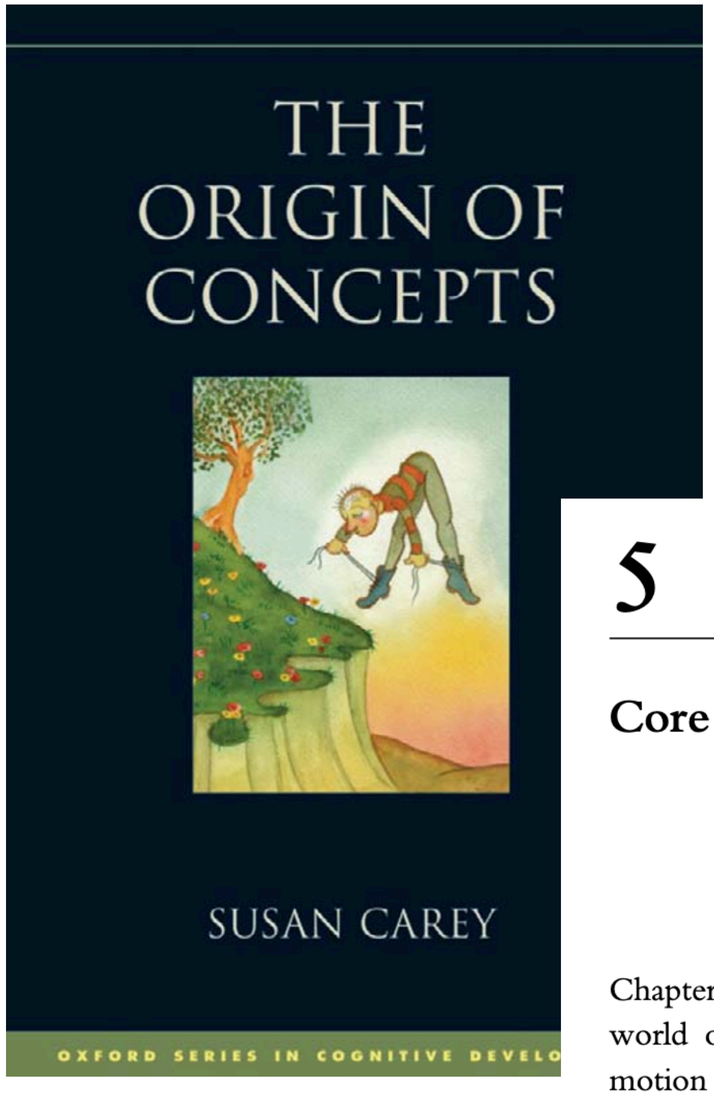
Minimal Mindreading

Emotion Attribution

Knowledge Attribution

Goal Attribution

Plan Attiribution



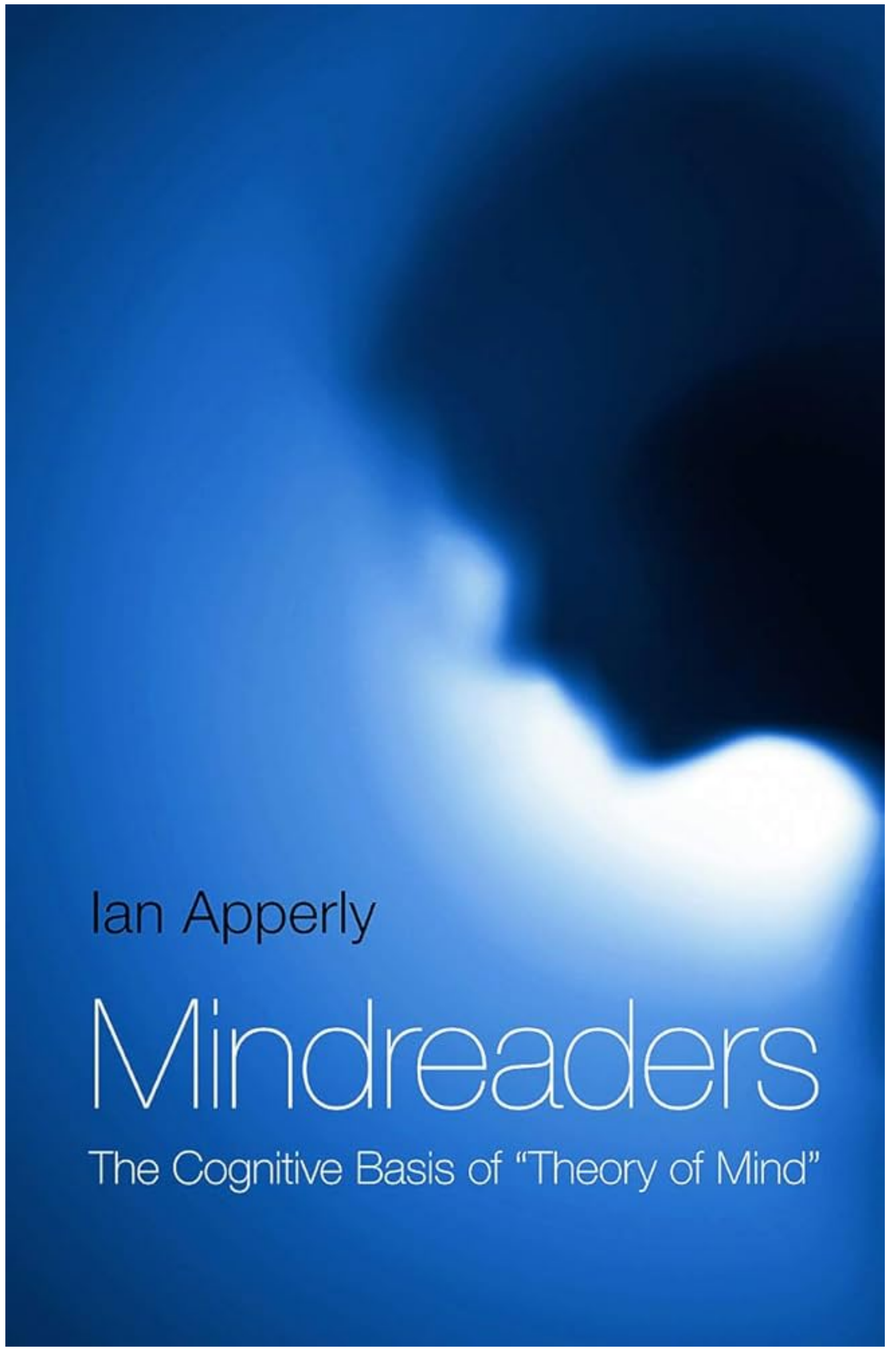
5

Core Cognition: Agency

Chapters 2 and 3 concerned core cognition of the physical world, the world of distinct individual objects and physical constraints on their motion and spatial relations. But the world of human infants is also social. For us primates especially, predicting what others of our kind will do, and influencing them to act in such a way that furthers our own interests, is crucial for our survival. Selection pressures on understanding others so as

Plan Attiribution

Goal Attribution



Belief Attribution

Minimal Mindreading

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

Behavioral and Brain Sciences

cambridge.org/bbs

Target Article

Cite this article: Phillips J, Buckwalter W, Cushman F, Friedman O, Martin A, Turri J, Santos L, Knobe J. (2021) Knowledge before belief. *Behavioral and Brain Sciences* **44**, e140: 1–75. doi:10.1017/S0140525X20000618

Target Article Accepted: 03 September 2020
Target Article Manuscript Online:

Knowledge before belief

Jonathan Phillips[a], Wesley Buckwalter[b], Fiery Cushman[c], Ori Friedman[d], Alia Martin[e], John Turri[f], Laurie Santos[g], and Joshua Knobe[h]

[a]Program in Cognitive Science, Department of Psychological and Brain Sciences and Department of Philosophy, Dartmouth College, Hanover, NH 03755, USA; [b]Department of Philosophy, Institute for Philosophy and Public Policy, George Mason University, Fairfax, VA 22030, USA; [c]Department of Psychology, Harvard University, Cambridge, MA 02138, USA; [d]Department of Psychology, University of Waterloo, Waterloo, ON, N2L 3G1, Canada; [e]School of Psychology, Victoria University of Wellington, Wellington, 6012, New Zealand; [f]Philosophy Department and Cognitive Science Program, University of Waterloo, Waterloo, ON N2L 3G1 Canada; [g]Department of Psychology, Yale University, New Haven, CT 06520, USA and [h]Program in Cognitive Science, Department of Philosophy, Yale University, New Haven, CT 06520, USA.
jonathan.s.phillips@dartmouth.edu; http://phillab.host.dartmouth.edu/
wesleybuckwalter@gmail.com; https://wesleybuckwalter.org/
cushman@fas.harvard.edu; http://cushmanlab.fas.harvard.edu/
friedman@uwaterloo.ca; https://sites.google.com/view/uwaterloocclab
alia.martin@vuw.ac.nz; https://vuwbabylab.com/
john.turri@gmail.com; https://john.turri.org/
laurie.santos@yale.edu; https://caplab.yale.edu/
joshua.knobe@yale.edu; https://campuspress.yale.edu/joshuaknobe/

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.

Assertoric mindreading

Peter van Elswyk
Northwestern University

This essay offers an explanation of how assertions express that the speaker has a propositional attitude towards what's asserted. The explanation is that this feature of assertion is owed to a hearer's spontaneous mindreading. I call this the assertoric mindreading hypothesis. Once developed and defended, the hypothesis is used to investigate which attitude is expressed. Since the attitude expressed is the attitude tracked during mindreading, the attitude must have a certain profile. It is argued that only factive attitudes like knowledge have this profile. Non-factive attitudes like belief or acceptance are ineligible.

Forthcoming in *Philosophy and Phenomenological Research*

Belief Attribution

Knowledge Attribution

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

Behavioral and Brain Sciences

cambridge.org/bbs

Target Article

Cite this article: Phillips J, Buckwalter W, Cushman F, Friedman O, Martin A, Turri J, Santos L, Knobe J. (2021) Knowledge before belief. *Behavioral and Brain Sciences* **44**, e140: 1–75. doi:10.1017/S0140525X20000618

Target Article Accepted: 03 September 2020
Target Article Manuscript Online:

Knowledge before belief

Jonathan Phillips[a], Wesley Buckwalter[b], Fiery Cushman[c], Ori Friedman[d], Alia Martin[e], John Turri[f], Laurie Santos[g], and Joshua Knobe[h]

[a]Program in Cognitive Science, Department of Psychological and Brain Sciences and Department of Philosophy, Dartmouth College, Hanover, NH 03755, USA; [b]Department of Philosophy, Institute for Philosophy and Public Policy, George Mason University, Fairfax, VA 22030, USA; [c]Department of Psychology, Harvard University, Cambridge, MA 02138, USA; [d]Department of Psychology, University of Waterloo, Waterloo, ON, N2L 3G1, Canada; [e]School of Psychology, Victoria University of Wellington, Wellington, 6012, New Zealand; [f]Philosophy Department and Cognitive Science Program, University of Waterloo, Waterloo, ON N2L 3G1 Canada; [g]Department of Psychology, Yale University, New Haven, CT 06520, USA and [h]Program in Cognitive Science, Department of Philosophy, Yale University, New Haven, CT 06520, USA.
jonathan.s.phillips@dartmouth.edu; http://phillab.host.dartmouth.edu/
wesleybuckwalter@gmail.com; https://wesleybuckwalter.org/
cushman@fas.harvard.edu; http://cushmanlab.fas.harvard.edu/
friedman@uwaterloo.ca; https://sites.google.com/view/uwaterloocclab
alia.martin@vuw.ac.nz; https://vuwbabylab.com/
john.turri@gmail.com; https://john.turri.org/
laurie.santos@yale.edu; https://caplab.yale.edu/
joshua.knobe@yale.edu; https://campuspress.yale.edu/joshuaknobe/

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.

Assertoric mindreading

Peter van Elswyk
Northwestern University

This essay offers an explanation of how assertions express that the speaker has a propositional attitude towards what's asserted. The explanation is that this feature of assertion is owed to a hearer's spontaneous mindreading. I call this the assertoric mindreading hypothesis. Once developed and defended, the hypothesis is used to investigate which attitude is expressed. Since the attitude expressed is the attitude tracked during mindreading, the attitude must have a certain profile. It is argued that only factive attitudes like knowledge have this profile. Non-factive attitudes like belief or acceptance are ineligible.

Forthcoming in *Philosophy and Phenomenological Research*

Belief Attribution

Knowledge Attribution

MUCH MORE ON FRIDAY!

Mindreading

Language

Belief Attribution

Minimal Mindreading

Emotion Attribution

Knowledge Attribution

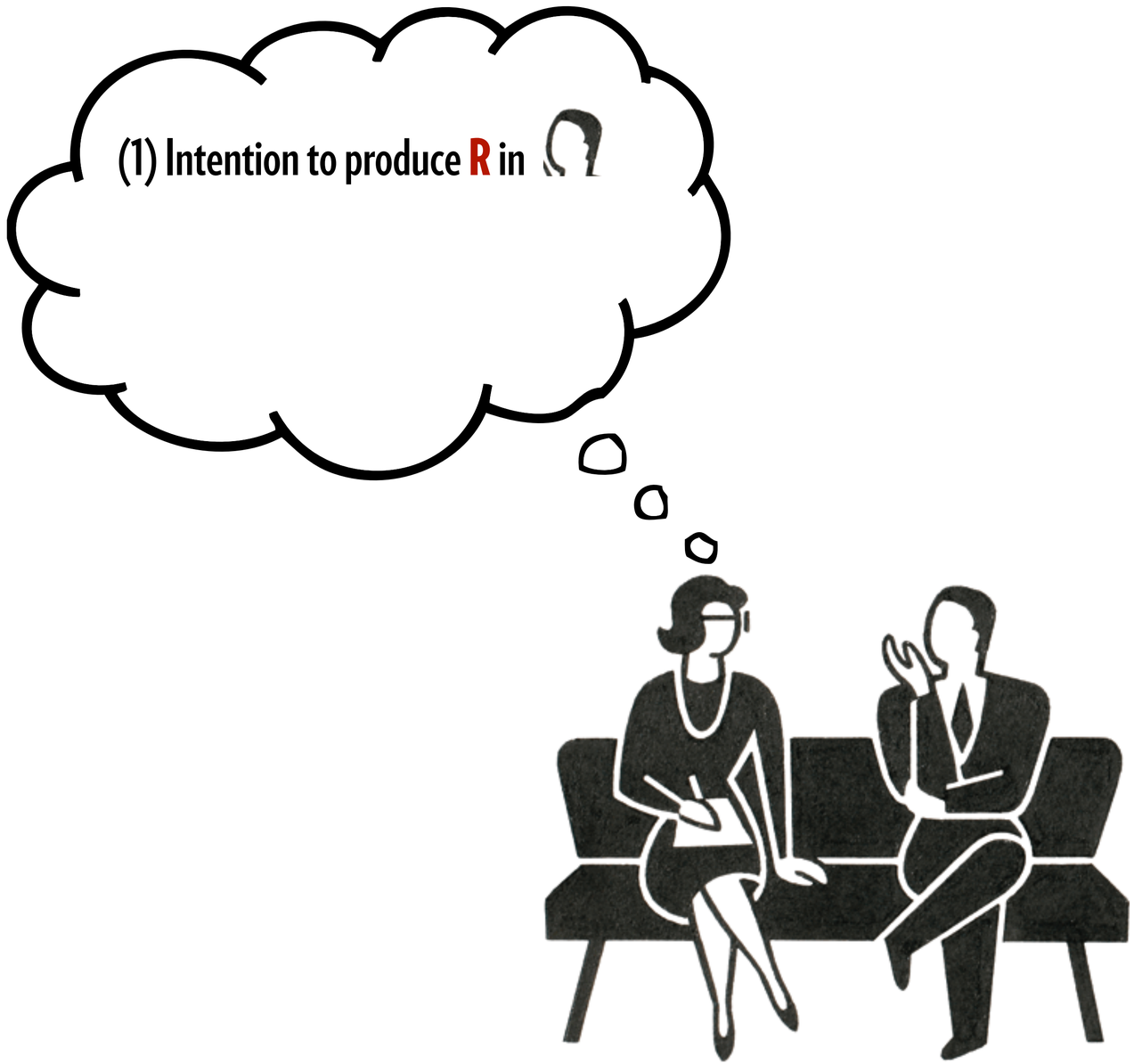
Goal Attribution

Plan Attiribution

SEMANTICS, PRAGMATICS, AND SOCIAL COGNITION

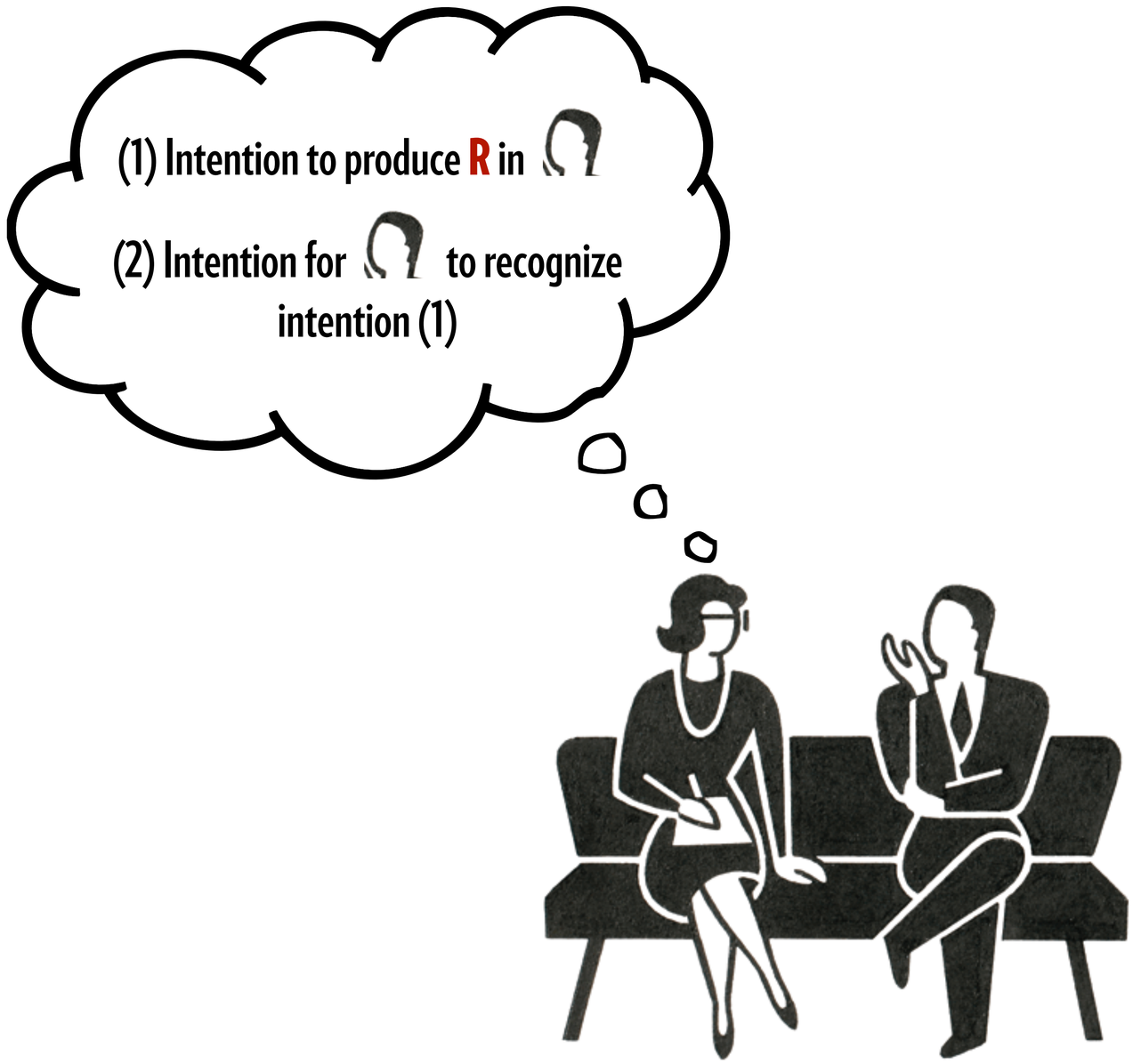
COMMUNICATION AS INTENTION RECOGNITION

(Grice 1957, 1969; etc.)



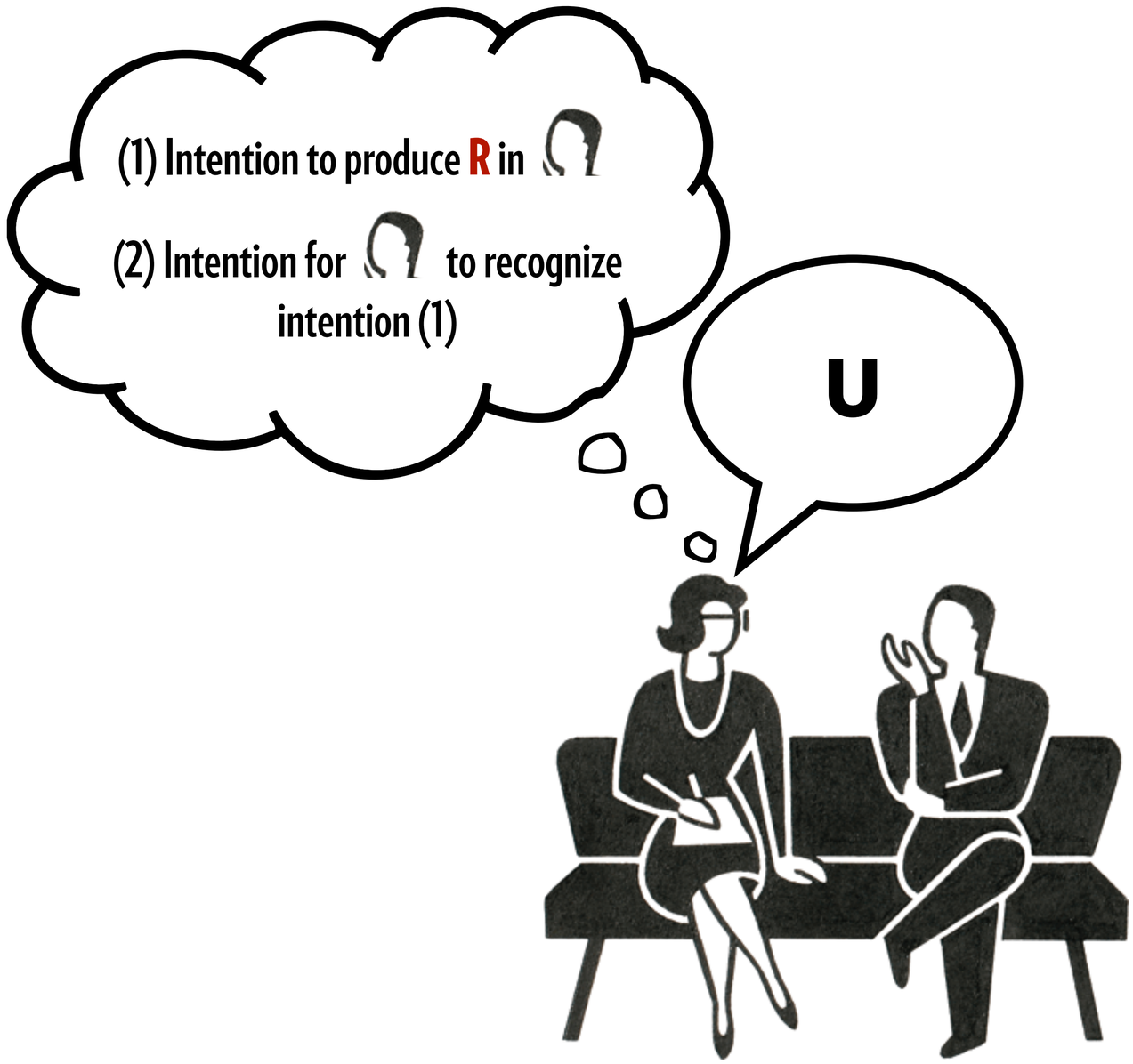
COMMUNICATION AS INTENTION RECOGNITION

(Grice 1957, 1969; etc.)



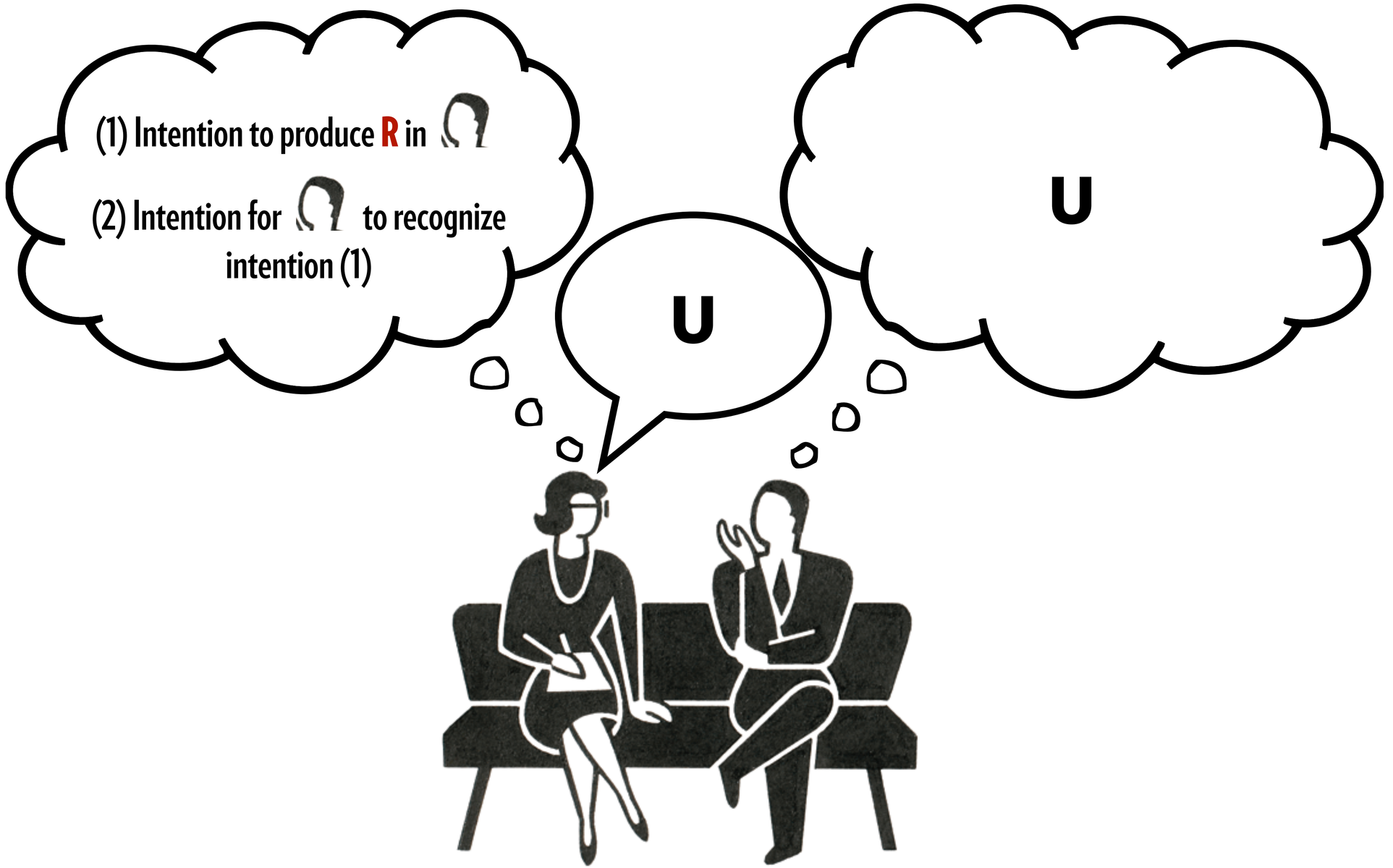
COMMUNICATION AS INTENTION RECOGNITION

(Grice 1957, 1969; etc.)



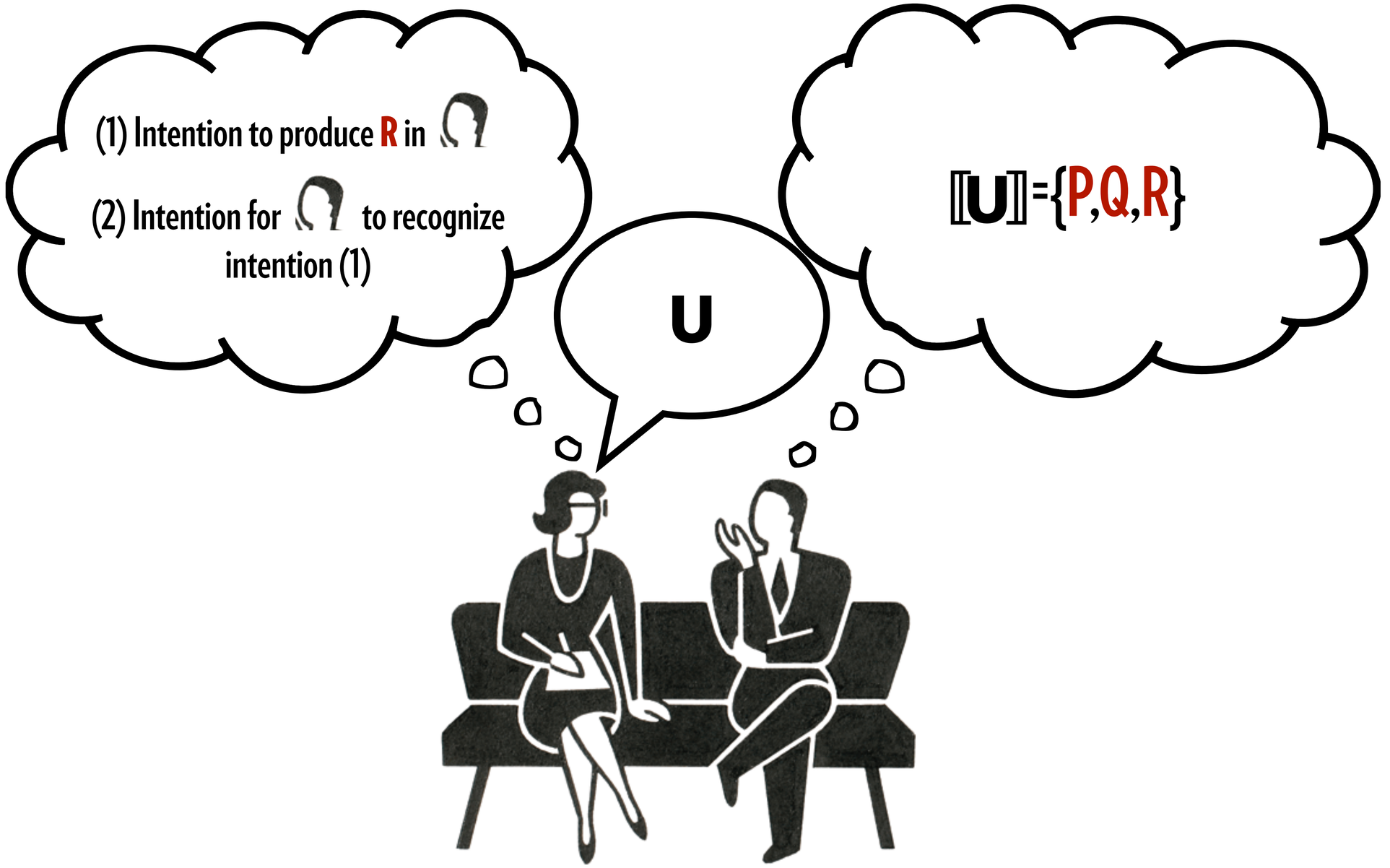
COMMUNICATION AS INTENTION RECOGNITION

(Grice 1957, 1969; etc.)



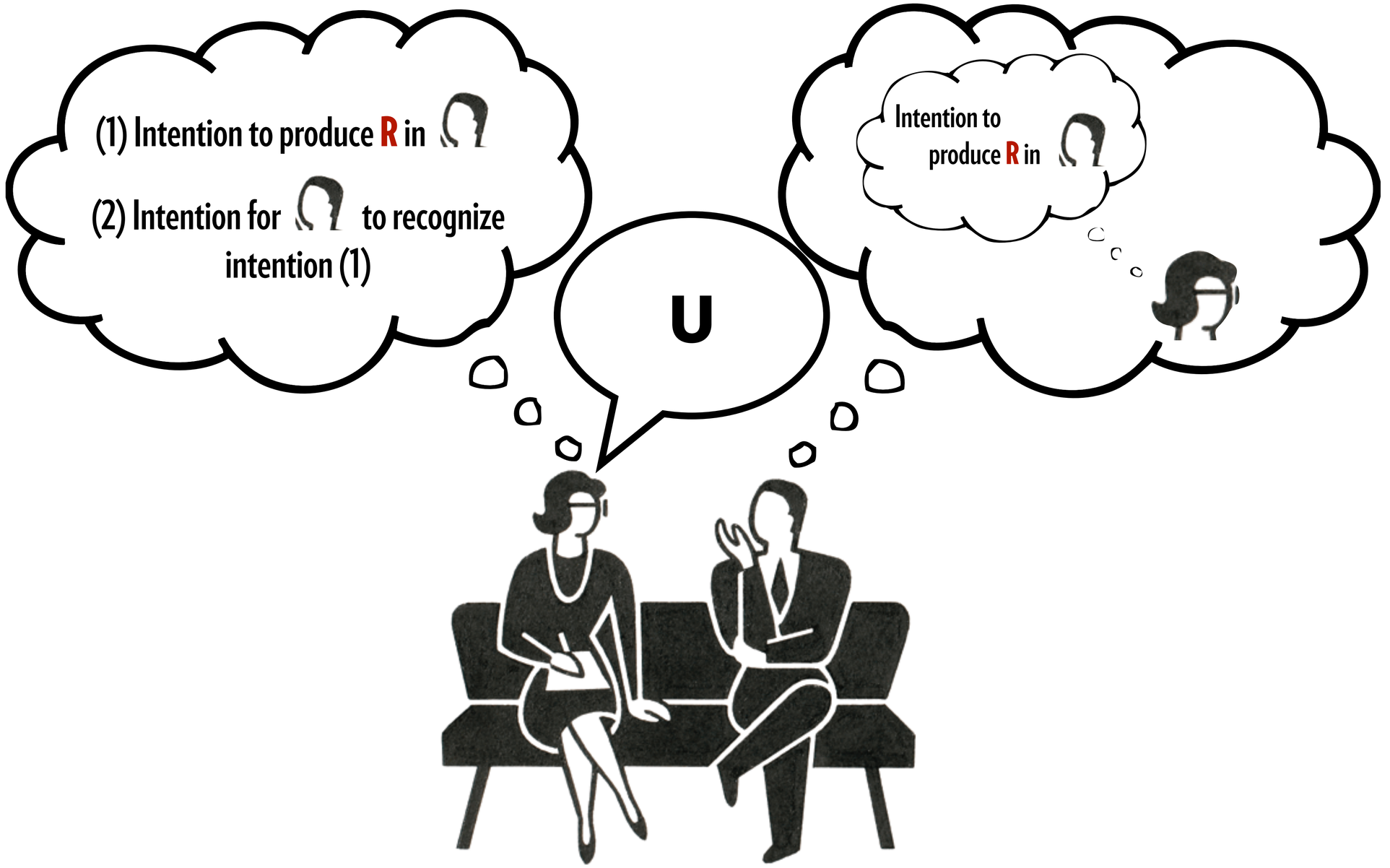
COMMUNICATION AS INTENTION RECOGNITION

(Grice 1957, 1969; etc.)



COMMUNICATION AS INTENTION RECOGNITION

(Grice 1957, 1969; etc.)

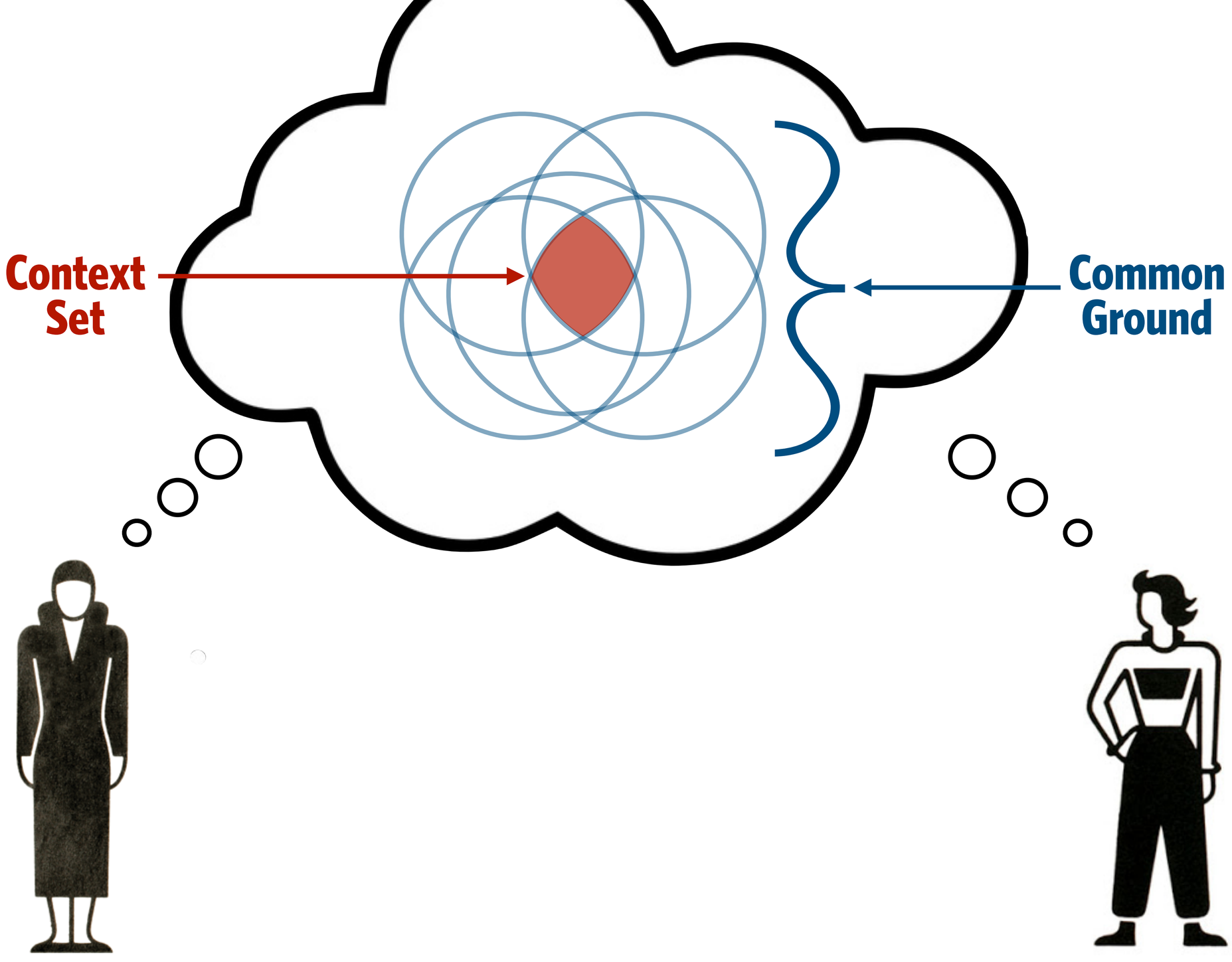


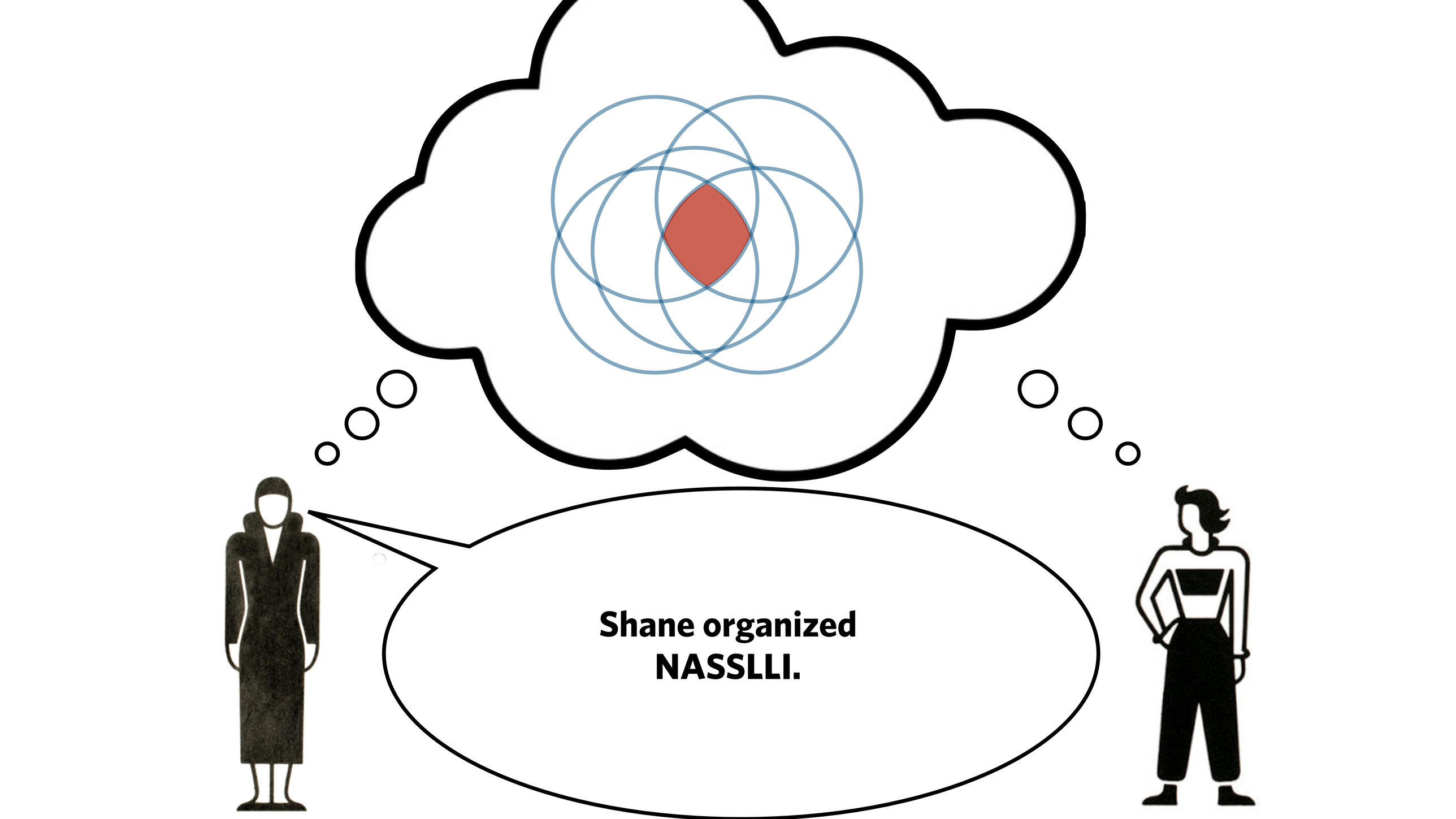
INTENTION RECOGNITION & SOCIAL COGNITION

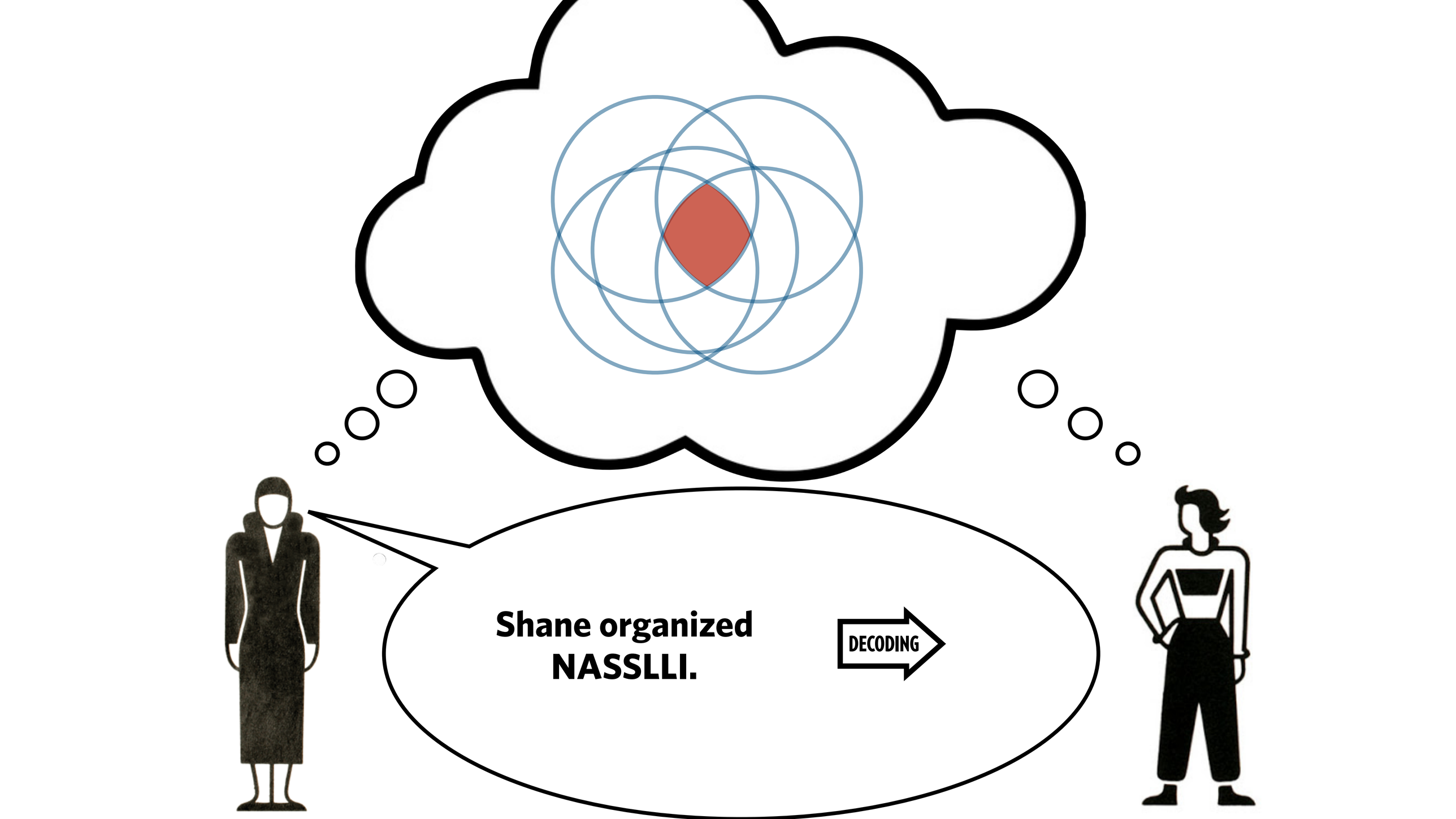
- **Message Design:** How do speakers decide what to communicate?
- **Signal Design:** How do speakers design utterances to give evidence of their intentions?
- How do hearers recognize speakers' intentions?
- Do we always communicate like this, or do we use shortcuts?
- If so, what kind of shortcuts?

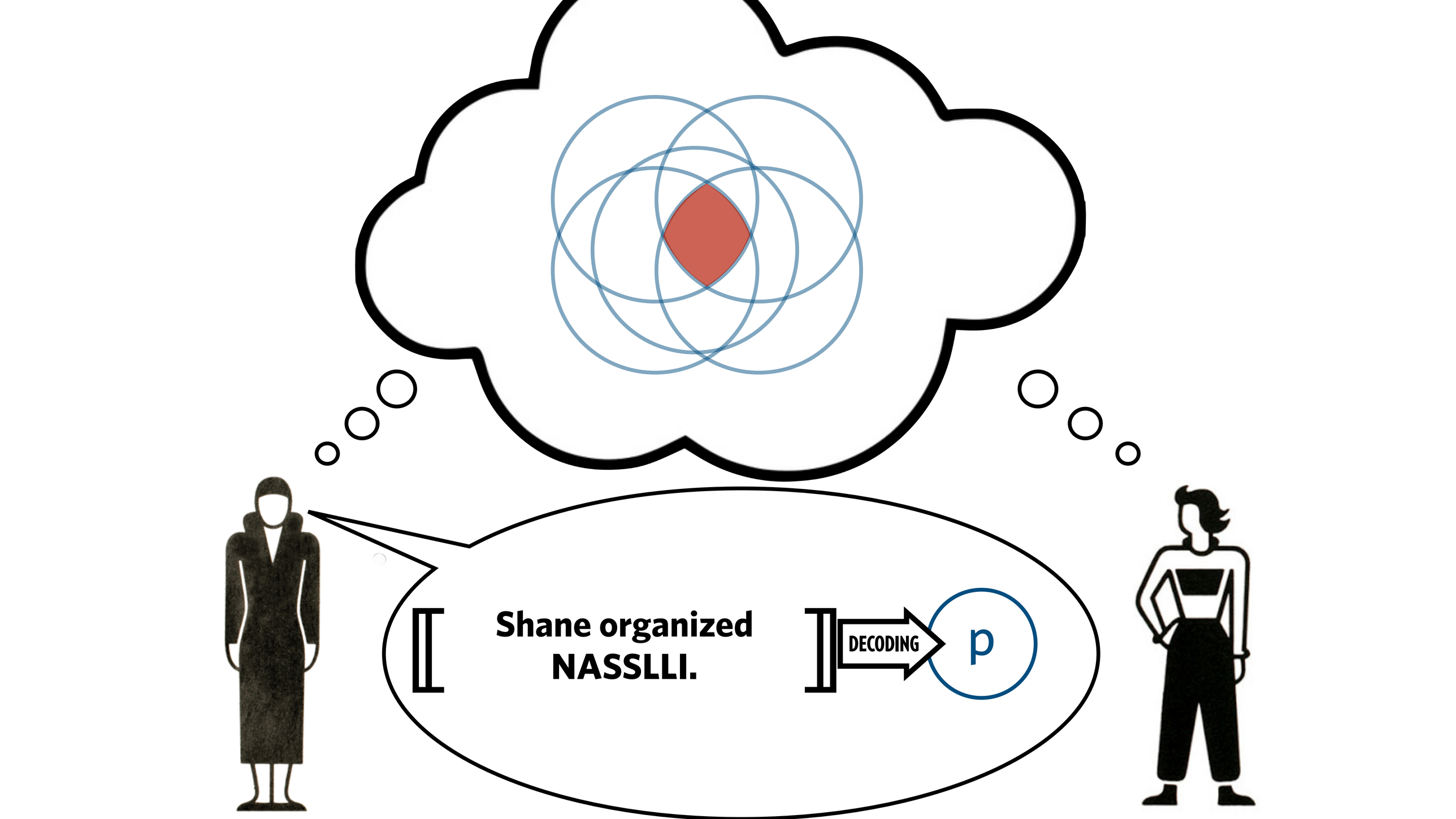
Common Ground

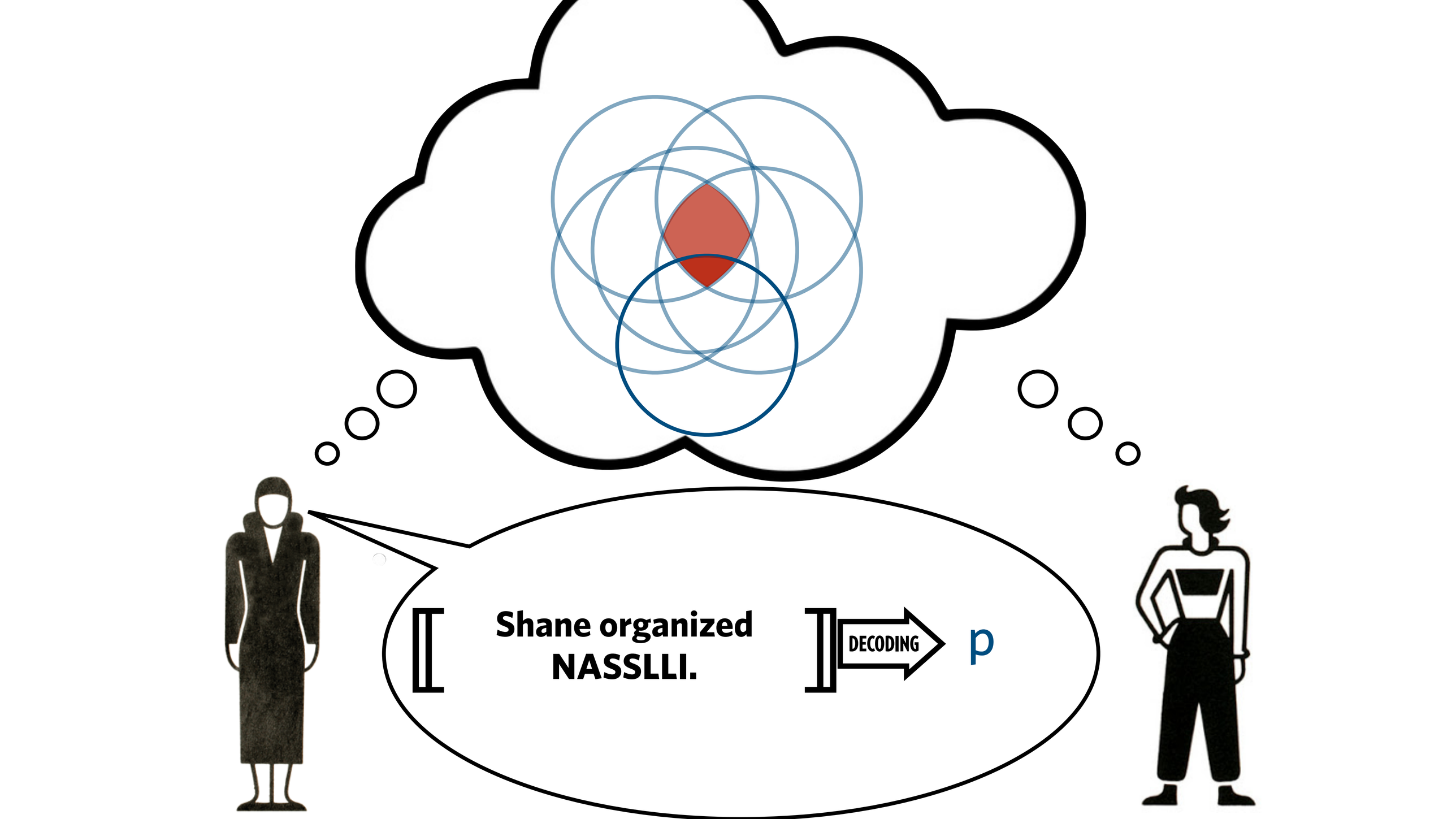
Stalnaker (1970, 1973, 1974, 1978, 2002, 2014);
Karttunnen (1974); Lewis (1979); etc.

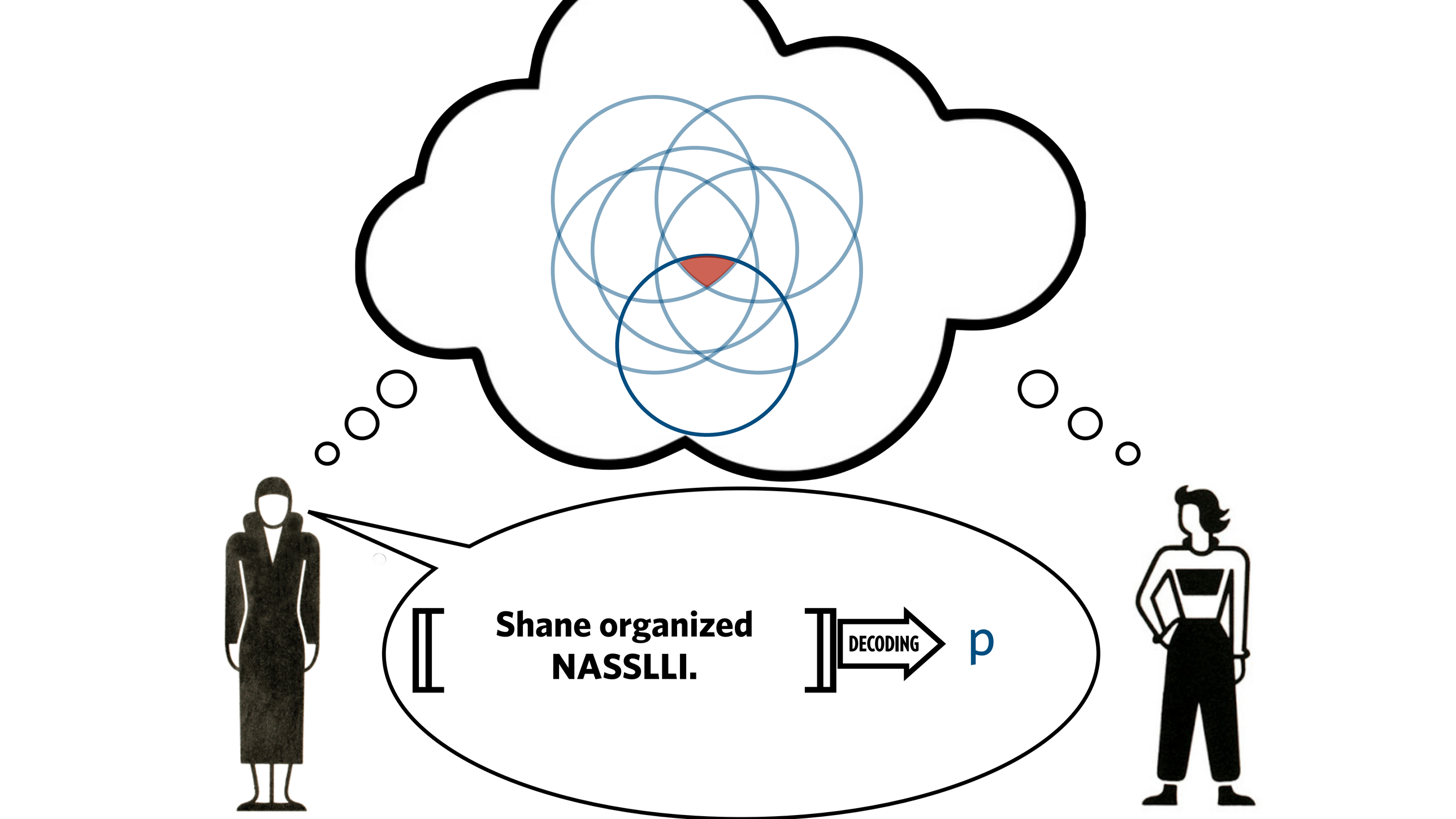

Context
Set
Common
Ground

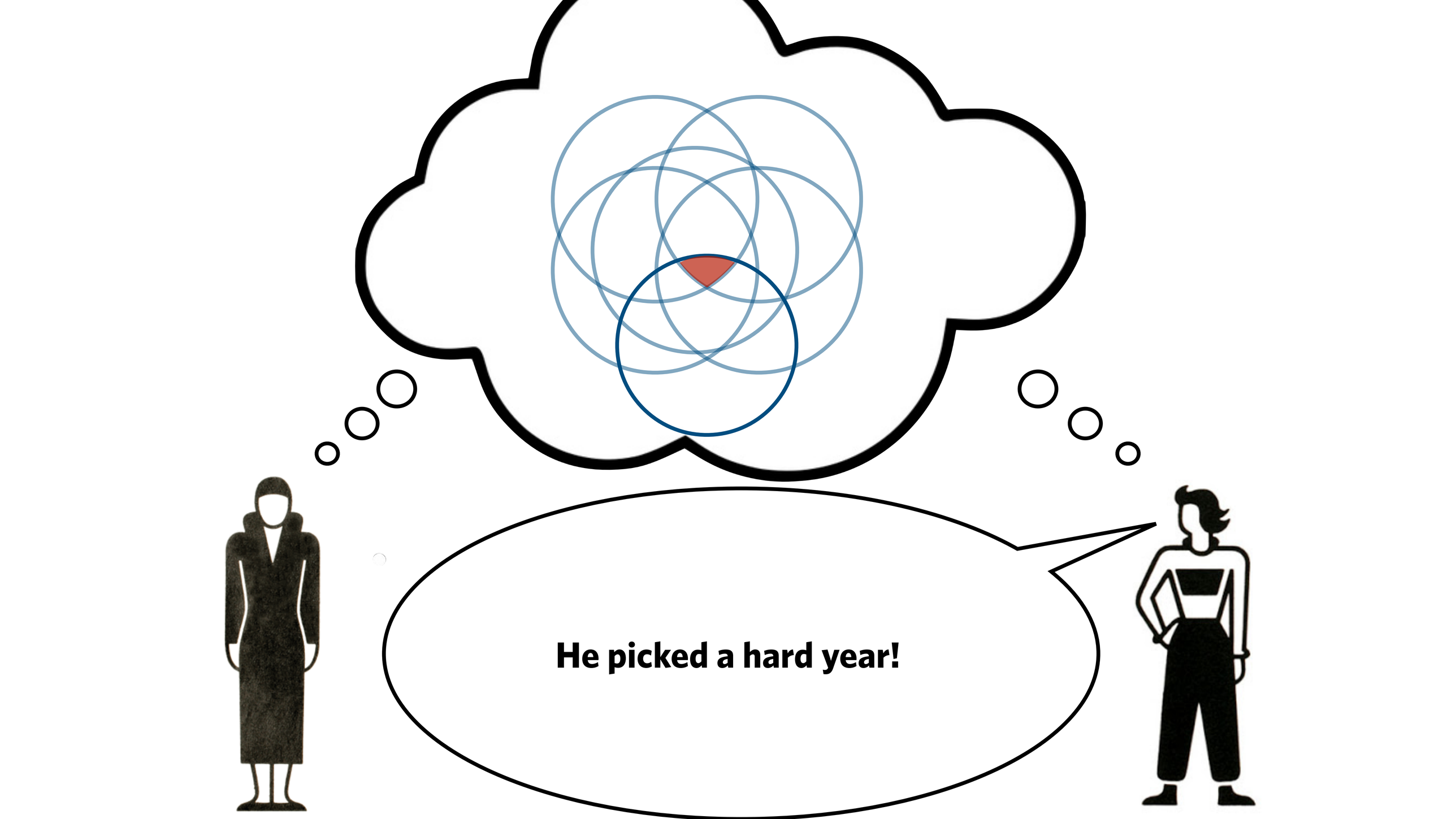

Shane organized
NASSLLI.

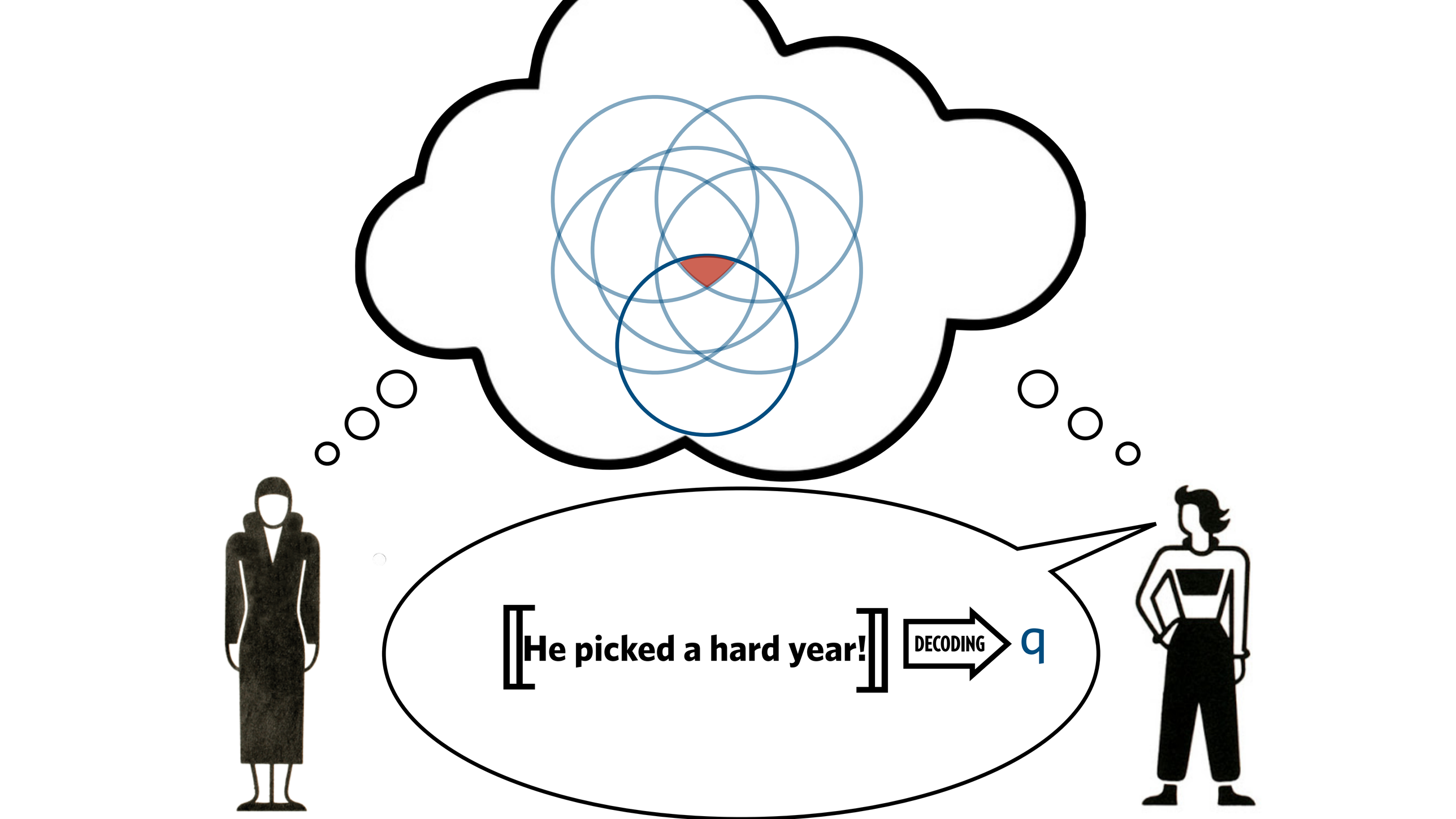

Shane organized NASSLLI.
DECODING

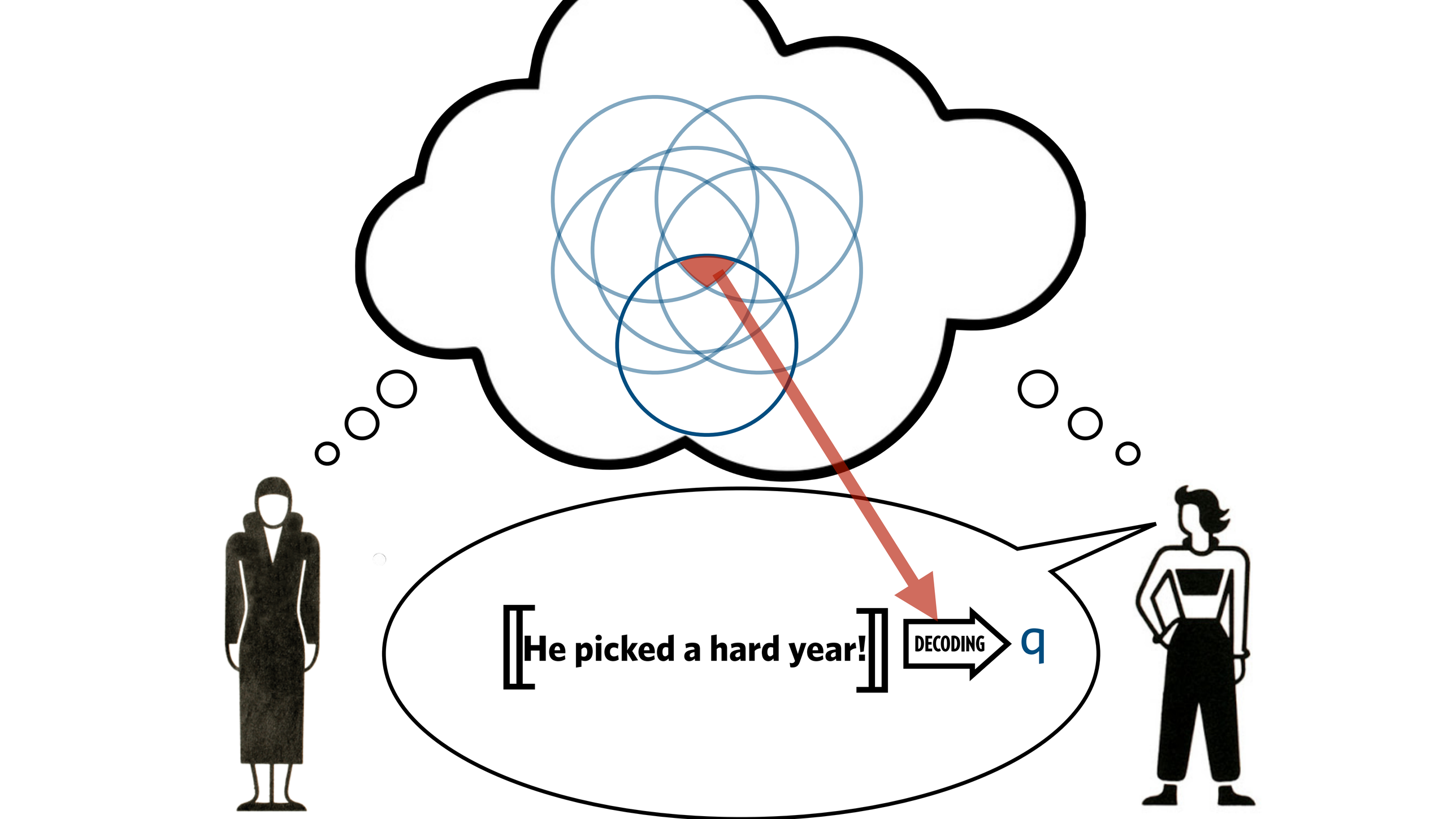

Shane organized NASSLLI.
DECODING
p


Shane organized NASSLLI.
DECODING
p


Shane organized NASSLLI.
DECODING
p


He picked a hard year!


[[He picked a hard year!]]
DECODING
q


[[He picked a hard year!]]
DECODING
q

Common Ground: Job Description

- Target of assertions
- Determines the contents of context-sensitive expressions
- Must satisfy (or accommodate) presuppositions
- Connects anaphora to antecedents
- Information relevant to audience design and interpretation

⋮

CG as Commonly Believed Joint Acceptance (Stalnaker 2002)

A proposition is CG for A and B (relative to some conversational purpose) iff:

1a. A accepts that p (for the purpose of the conversation);

1b. B accepts that p (for the purpose of the conversation);

2a. A believes (1a–b);

2b. B believes (1a–b);

3a. A believes that B believes (1a–b);

3b. B accepts that A believes (1a–b);

⋮

Common Ground and Social Cognition: Questions

- Can we really have infinite, intersubjectively iterated propositional attitudes?
- What cognitive mechanisms do we use to get, maintain, and coordinate those attitudes?
- Always the same mechanisms?
- What kind of cognitive resources does this take?
- What about kids and animals?
- ⋮

Common Ground and Social Cognition: Questions

- Can we really have infinite, intersubjectively iterated propositional attitudes?
- What cognitive mechanisms do we use to get, maintain, and coordinate those attitudes?
- Always the same mechanisms?
- What kind of cognitive resources does this take?
- What about kids and animals?
- ⋮

MORE ON THURSDAY!

Evidentials: Brief description

- Evidentials are words or morphemes that indicate an agent's *source* of evidence.
- Sources are indicated to be *direct* or *indirect*, or a subtype (e.g. inferred, reported).
- Evidentials can embed in matrix declaratives and sometimes polar interrogatives.
- Evidentials can be grammatically obligatory or optional in these environments.

Evidentials: Turkish illustration

- Turkish has two evidentials: *-dı* (direct) and *-mış* (indirect).
- Turkish evidential are obligatory in the past tense for declaratives and polar interrogatives.

Vazo kırıl dı.	The vase was broken, I experienced.
Vazo kırıl mış.	The base was broken, I learned.

Evidentials: Open questions

- How are evidentials related to propositional attitudes? (Hearsay evidential are acceptable with a denial of belief akin to *p, I heard, but I don't believe that* (AnderBois 2014), but direct evidentials are not (Murray (2017)).

- How do evidentialized declaratives update the common ground? (Is it a secondary proposition (Murray 2017), are there multiple common grounds (Northrup 2014), or what?)

Evidentials: Open questions

How the open questions are answered makes predictions for how language interfaces with social cognition.

Option #1
Evidentials *do* encode the speaker's attitude

Prediction
The acquisition of evidentials is not independent of the emergence of mindreading in children.

Option #2
Evidentials *do not* encode the speaker's attitude

Prediction
The acquisition of evidentials is independent of the emergence of mindreading in children.

Evidentials: Open questions

JOURNAL OF COGNITION AND DEVELOPMENT
https://doi.org/10.1080/15248372.2023.2260874

Routledge
Taylor & Francis Group

Check for updates

What's the Evidence Say? The Relation Between Evidential-Trust and Theory of Mind

Bartuğ Çelik[a], Nice Ergut[b], and Jedediah W.P. Allen[c]

[a]Central European University, Austria; [b]TED University, Turkey; [c]Bilkent University, Turkey

ABSTRACT

Previous research has shown that linguistic cues such as mental and modal verbs can influence young children's judgments about the reliability of informants. Further, certain languages include grammatical morphemes (i.e. evidential markers), which clarify the source of information coming from testimony (e.g., Bulgarian, Japanese, Turkish). Accordingly, the first aim of the current study is to examine whether Turkish-speaking children's reliability judgments change based on evidential markers (i.e. the past-tense *direct* evidential marker, *–DI*, and the past-tense *indirect* evidential marker, *–mIş*). The literature has also investigated whether selective trust abilities are related to understanding the epistemic states of others (i.e., Theory of Mind). Therefore, the second aim is to examine the relation between selective trust based on evidential markers and ToM abilities by including a comprehensive ToM battery. Eighty-six Turkish-speaking preschool and elementary school children between the ages of 4 and 7, residing in a metropolitan city in Turkey, participated in a selective trust task based on evidential markers, a general language task, and a ToM battery. The results of the current study showed that after the age of 6, Turkish-speaking children start to selectively trust the informant using the past-tense direct evidential marker, *–DI*, over the past-tense indirect evidential marker, *–mIş*. Selective trust performance was related to receptive vocabulary but not to ToM abilities after controlling for participant gender and age. Overall, the results contribute to current discussions about children's selective trust ability based on linguistic cues and its relation with ToM.

KEYWORDS
selective trust; theory of mind; evidential markers; reliability; Turkish

Due to its opaqueness, children need reliable testimony to learn cultural knowledge about many aspects of the physical and social world. Learning about sources of testimony is thus necessary to grasp whether such information is reliable or not. In order to explore the reliability of testimony, questions such as "Based on what?" or "Where did you get your information?" could be asked. The informant's reply could be based on *direct* (first-hand) experience such as "I saw the event," or *indirect* (second-hand) experience, "I heard it from someone." Research has shown that by the age of 3, children's judgments about the reliability of informants vary based on (non-linguistic) epistemic cues such as past accuracy (Jaswal & Neely, 2006; Koenig & Harris, 2005; Pasquini, Corriveau, Koenig, & Harris, 2007)

Received: 2 September 2022 | Revised: 16 January 2023 | Accepted: 16 March 2023

DOI: 10.1111/desc.13398

Developmental Science

WILEY

RESEARCH ARTICLE

Culturally constituted universals: Evidential basis of belief matters

Feride Nur Haskaraca[1] | **Hande Ilgaz**[2]

[1]Department of Developmental Psychology, University of Göttingen, Göttingen, Germany

[2]Psychology Department, Bilkent University, Ankara, Turkey

Correspondence
Hande Ilgaz, Bilkent University, Psychology Department, 06800 Ankara, Turkey.
Email: hande.ilgaz@bilkent.edu.tr

Abstract

Differences in the sequence with which children pass the tasks in Wellman and Liu's (2004) theory of mind (ToM) battery is increasingly bringing into question the universal and cultural specifics of children's developing understanding of others' minds. Children from China, Iran, and Turkey pass the knowledge access (KA) task of the battery earlier than they pass the diverse beliefs (DB) task (e.g., Selcuk et al., 2018). This pattern is the reverse of what has been documented with children from Australia and the US (e.g., Peterson et al., 2005). This paper presents three studies with Turkish samples that explore the possible reasons for developmental sequence and performance differences in the ToM battery. Study 1 investigated Turkish-speaking adults' judgments of appropriateness for different epistemic verbs as used in the DB and false belief (FB) tasks. Study 2 investigated whether adults' performance (i.e., accuracy, reaction time) on FB tasks were affected by culturally preferred uses of these verbs. Collectively these studies showed that adults found different epistemic language ("guess," "think," and the Turkish-specific "falsely think") to be appropriate for different belief-based tasks. However, there was no difference in adults' performance based on epistemic language. In Study 3, Turkish-speaking preschoolers' performance in belief-related tasks based on variations in epistemic language and epistemic features (i.e., presence of evidence) was investigated. Among five modifications, Turkish children benefited only from a modification that involved the manipulation of the epistemological basis for ambiguous beliefs (i.e., visual evidence for belief).

KEYWORDS
cross-cultural variation, diverse beliefs, evidentiality, theory of mind, theory of mind scale

THE RATIONAL SPEECH ACT (RSA) MODEL

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.

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.

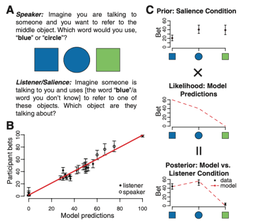


Fig. 1. (A) An example stimulus from our experiment, with instructions for speaker, listener, and salience conditions. **(B)** Human bets on the probability of a choosing a term (speaker condition, $N = 206$) or referring to an object (listener condition, $N = 263$), plotted by model predictions. Points 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.

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 Macmillian, 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.

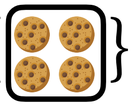
u_{none} = "Alex ate none of the cookies"

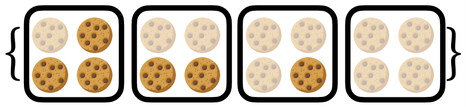
u_{some} = "Alex ate some of the cookies"

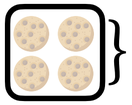
u_{all} = "Alex ate all of the cookies"

Semantics

The literal meanings of the utterances in U.

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

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

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

Basic Scalar Implicature Game

$P(m_0) = 0.2$ 

$P(m_1) = 0.2$ 

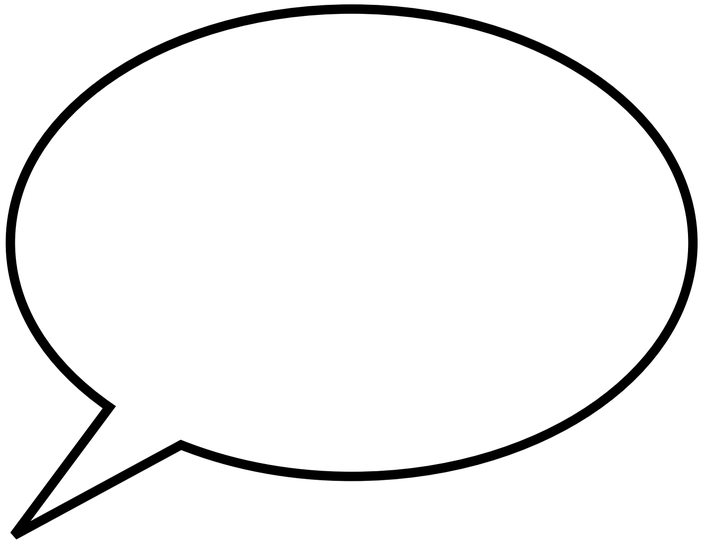
$P(m_2) = 0.2$ 

$P(m_3) = 0.2$ 

$P(m_4) = 0.2$ 

Basic Scalar Implicature Game

- 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.

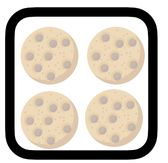
$P(m_0|u_{all}) = 0$

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

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

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

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



Basic Scalar Implicature Game

- Similar story here.

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

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

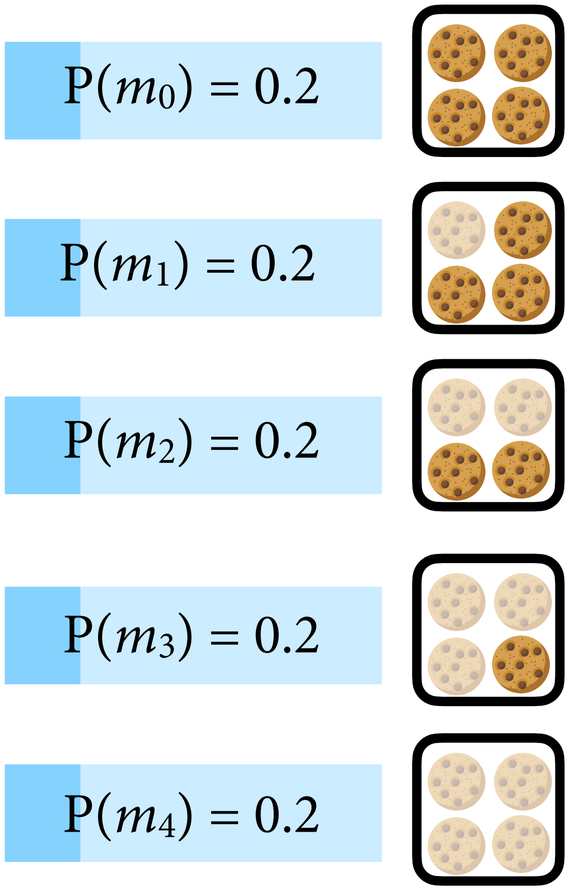
$P(m_2|u_{none}) = 0$

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

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

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?



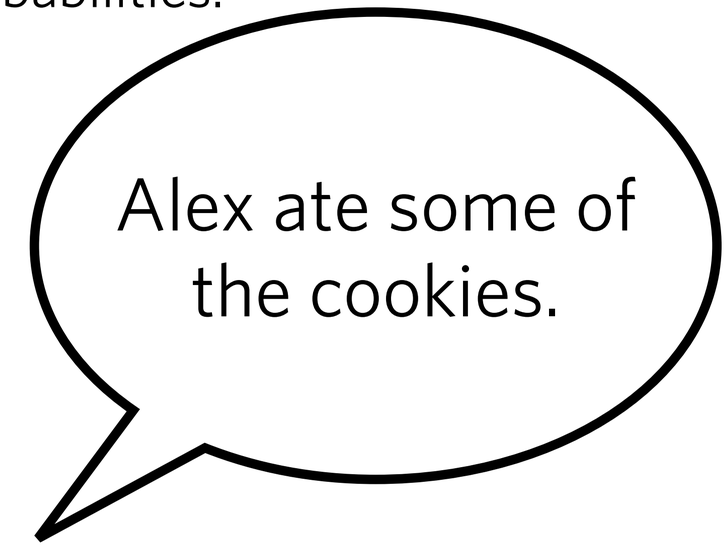
Basic Scalar Implicature Game

- 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 alternatives compatible with the literal meaning of u, in proportion to their prior probabilities.



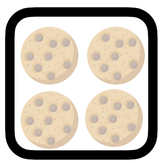
$P_{L0}(m_0|u_{some}) = 0$

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

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

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

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



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.

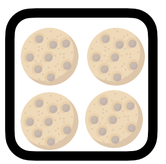
$P_{L0}(m_0|u_{some}) = 0$

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

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

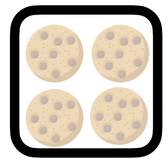
$P_{L0}(m_3|u_{some}) = .25$

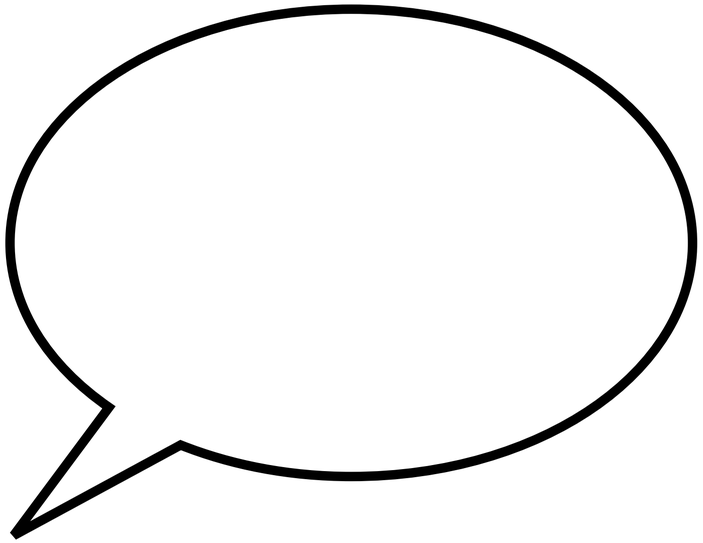
$P_{L0}(m_4|u_{some}) = .25$



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.





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

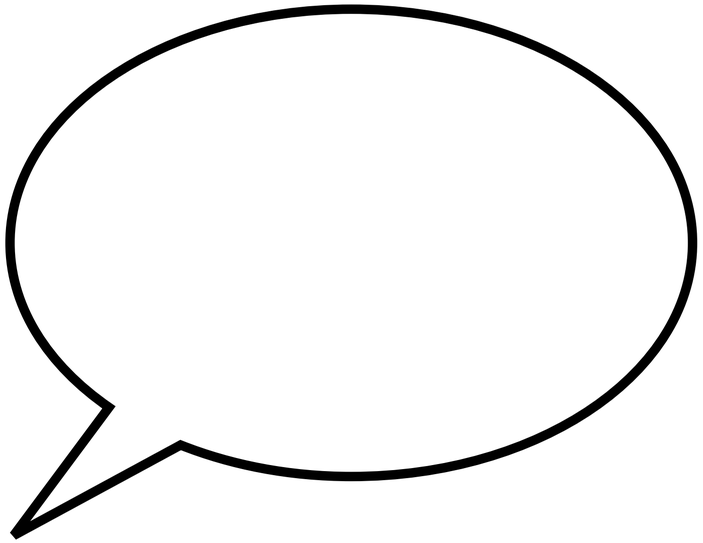
$$P(u_{\text{some}} | m_{1\text{–}3}) = ?$$

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

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:

$$\mathrm{U}(u,m) = \ln \mathrm{P}_{\mathrm{L0}}(m|u) - \mathrm{cost}(u)$$



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

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

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

Basic Scalar Implicature Game

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

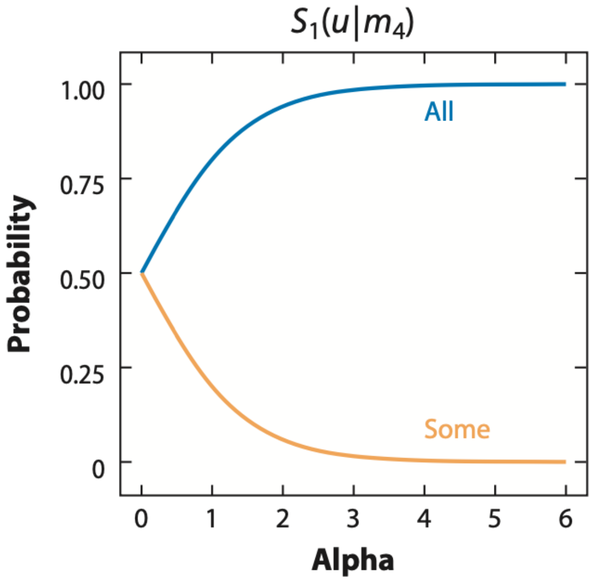
$$P_{S1}(u|m) \propto \exp(\alpha \cdot U(u; m))$$

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

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

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

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



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

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

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

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.

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

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

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

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

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

RSA and Social Cognition: Questions

- Does it follow from RSA models that we have to represent and anticipate other agents' partial beliefs (credences), and not merely their full beliefs?
- How do we do this?
- With what degree of precision?
- How hard is it?
- ⋮

RSA and Social Cognition: Questions

- Does it follow from RSA models that we have to represent and anticipate other agents' partial beliefs (credences), and not merely their full beliefs?
- How do we do this?
- With what degree of precision?
- How hard is it?
- ⋮

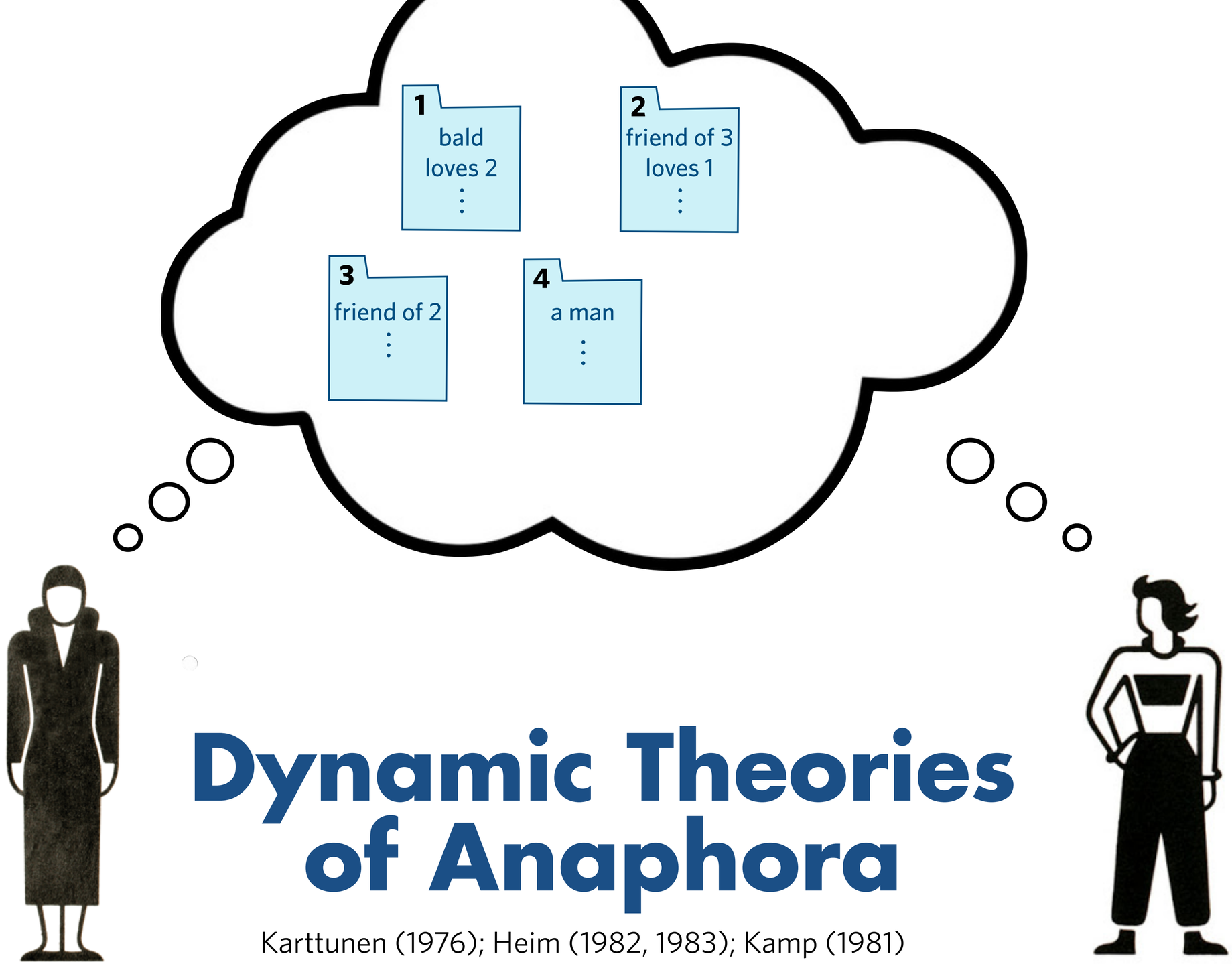
SCIENCE ADVANCES | RESEARCH ARTICLE

PSYCHOLOGICAL SCIENCE

Quantitative mental state attributions in language understanding

Julian Jara-Ettinger[1,2]* and Paula Rubio-Fernandez[3,4]

Human social intelligence relies on our ability to infer other people's mental states such as their beliefs, desires, and intentions. While people are proficient at mental state inference from physical action, it is unknown whether people can make inferences of comparable granularity from simple linguistic events. Here, we show that people can make quantitative mental state attributions from simple referential expressions, replicating the fine-grained inferential structure characteristic of nonlinguistic theory of mind. Moreover, people quantitatively adjust these



Dynamic Theories of Anaphora

Karttunen (1976); Heim (1982, 1983); Kamp (1981)

One of the ten marbles is not in the bag. It is probably under the sofa.

Nine of the ten marbles are in the bag. It is probably under the sofa.

Partee/Heim

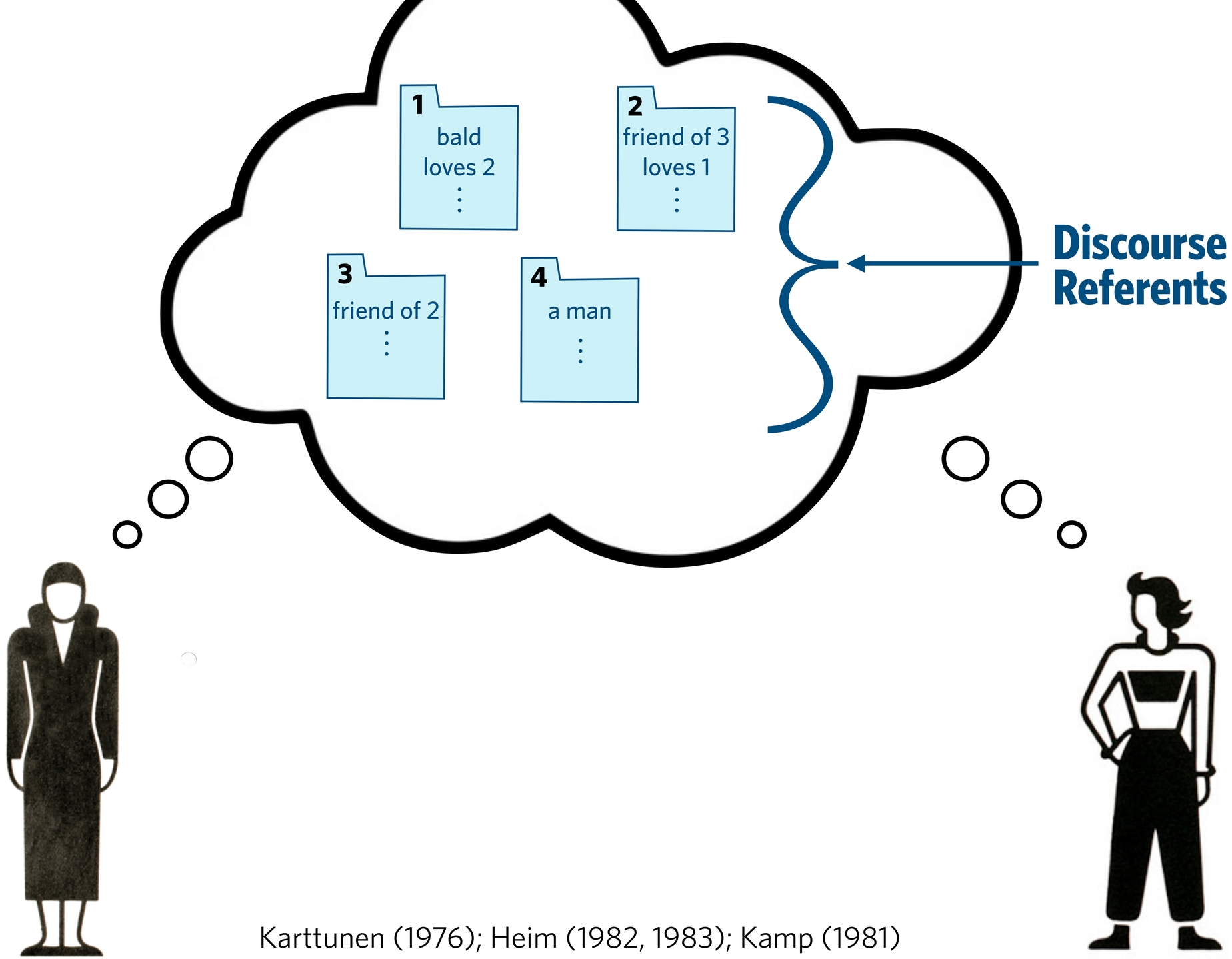
He is coming to the party tonight.

I know a man.

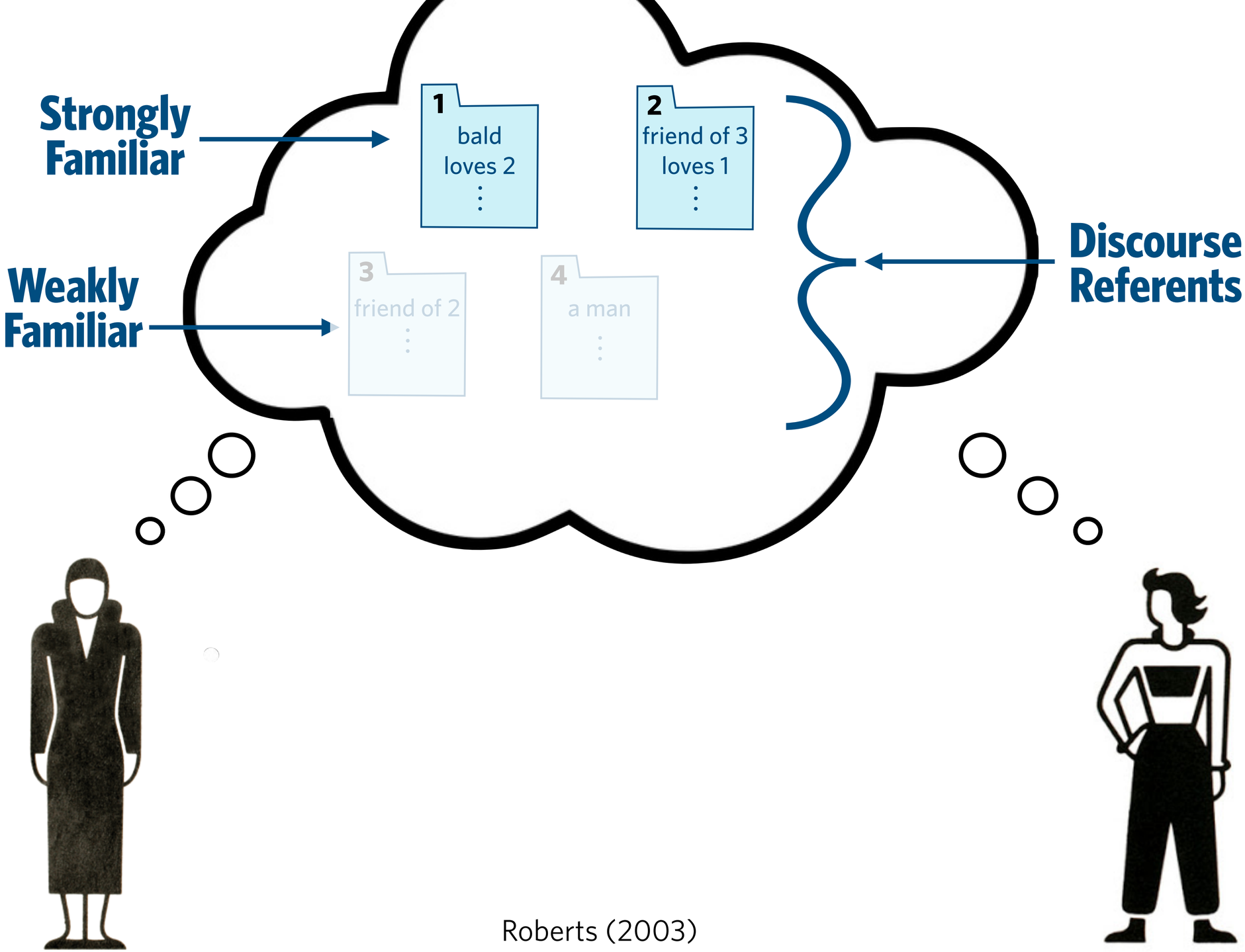
He is coming to the party tonight.

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.

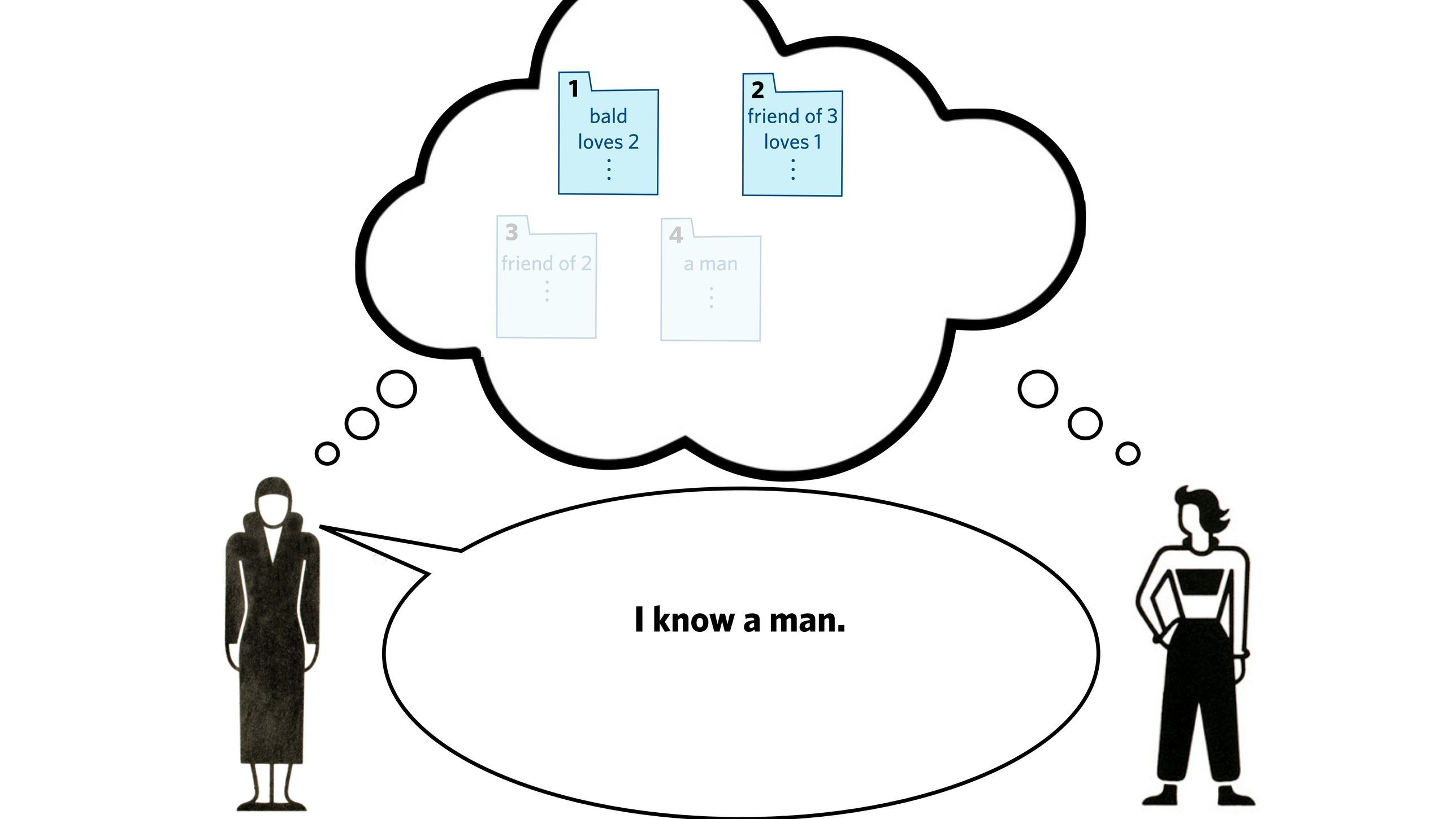
—Heim (1982)

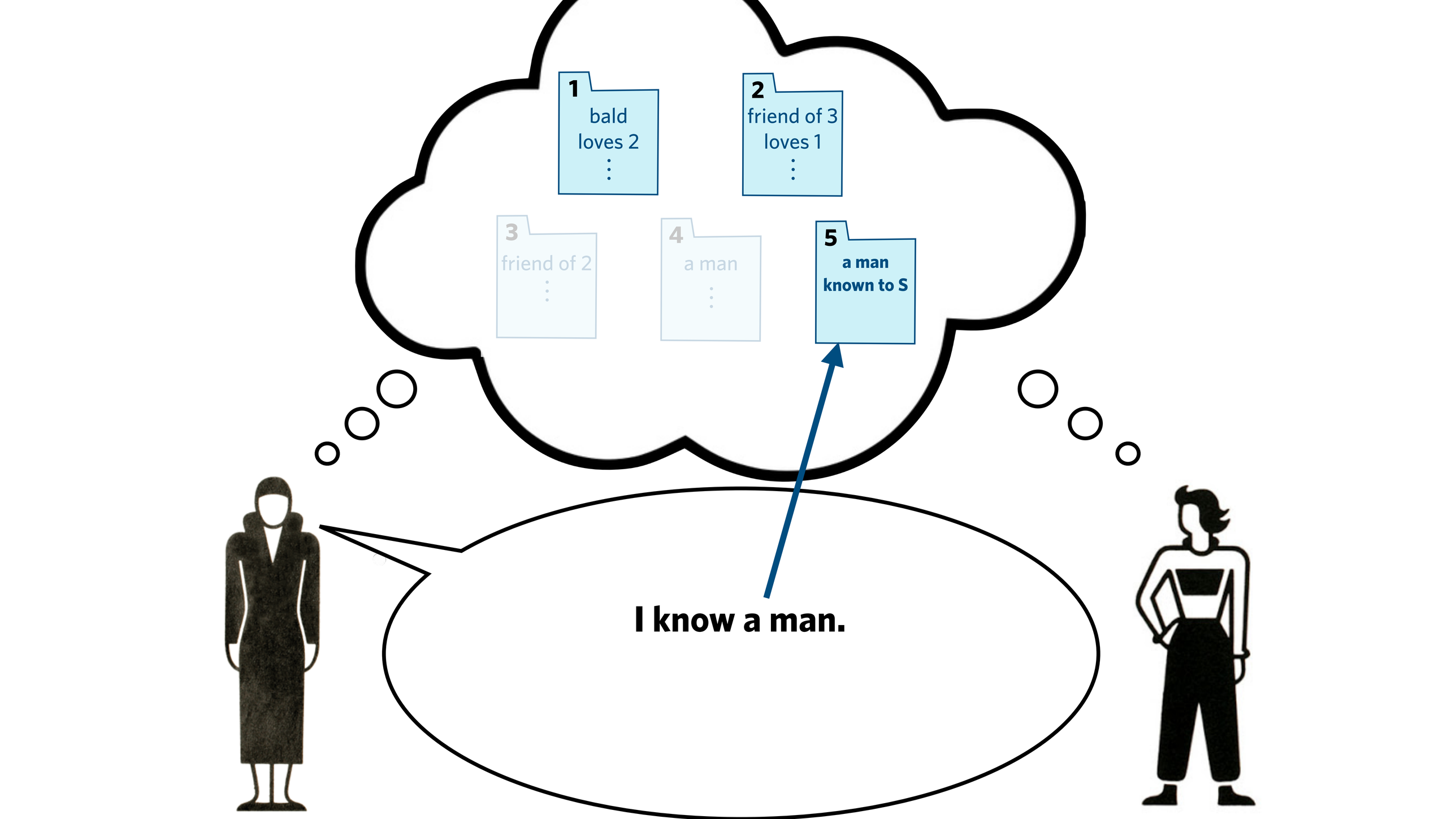


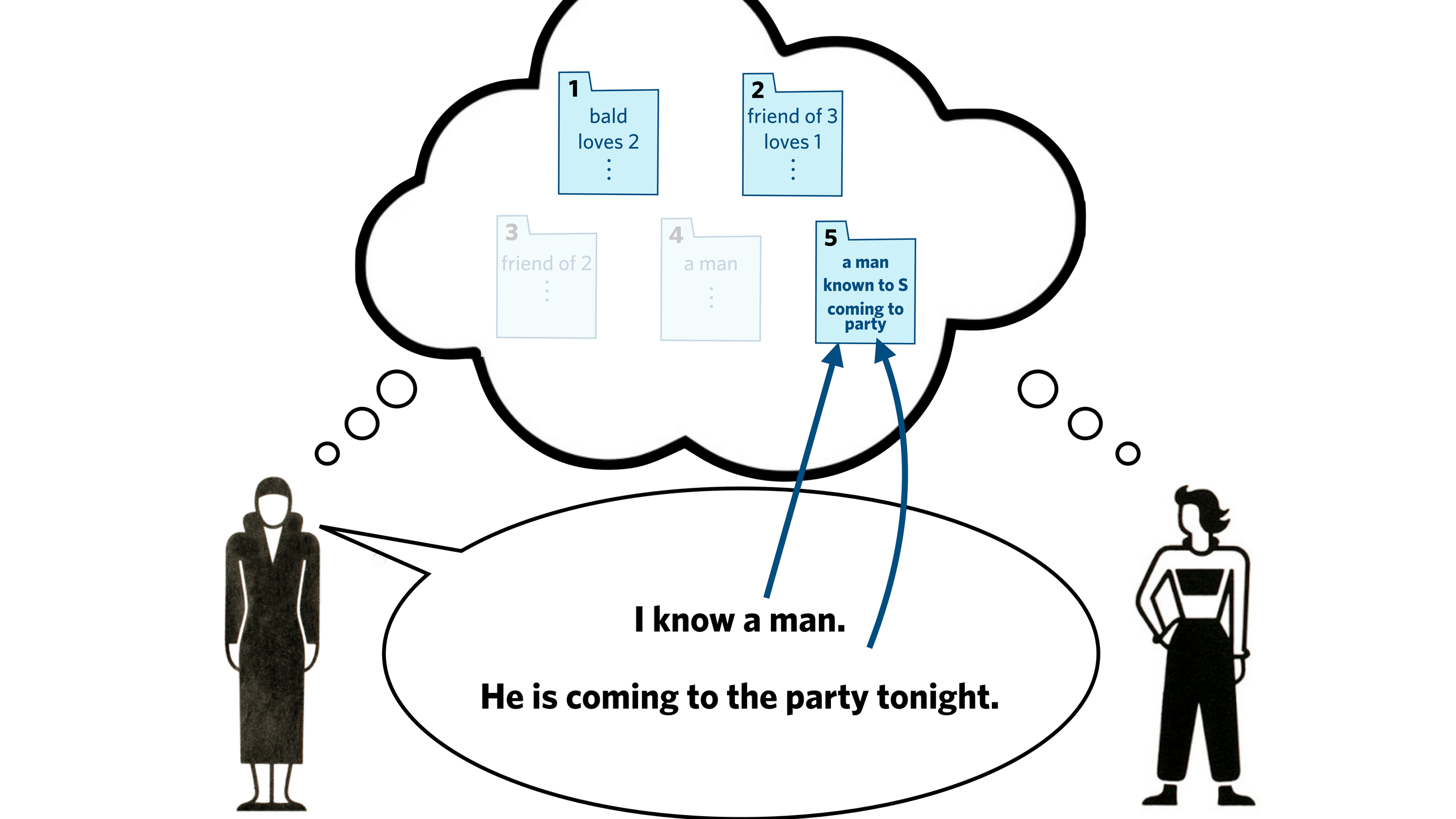
Karttunen (1976); Heim (1982, 1983); Kamp (1981)

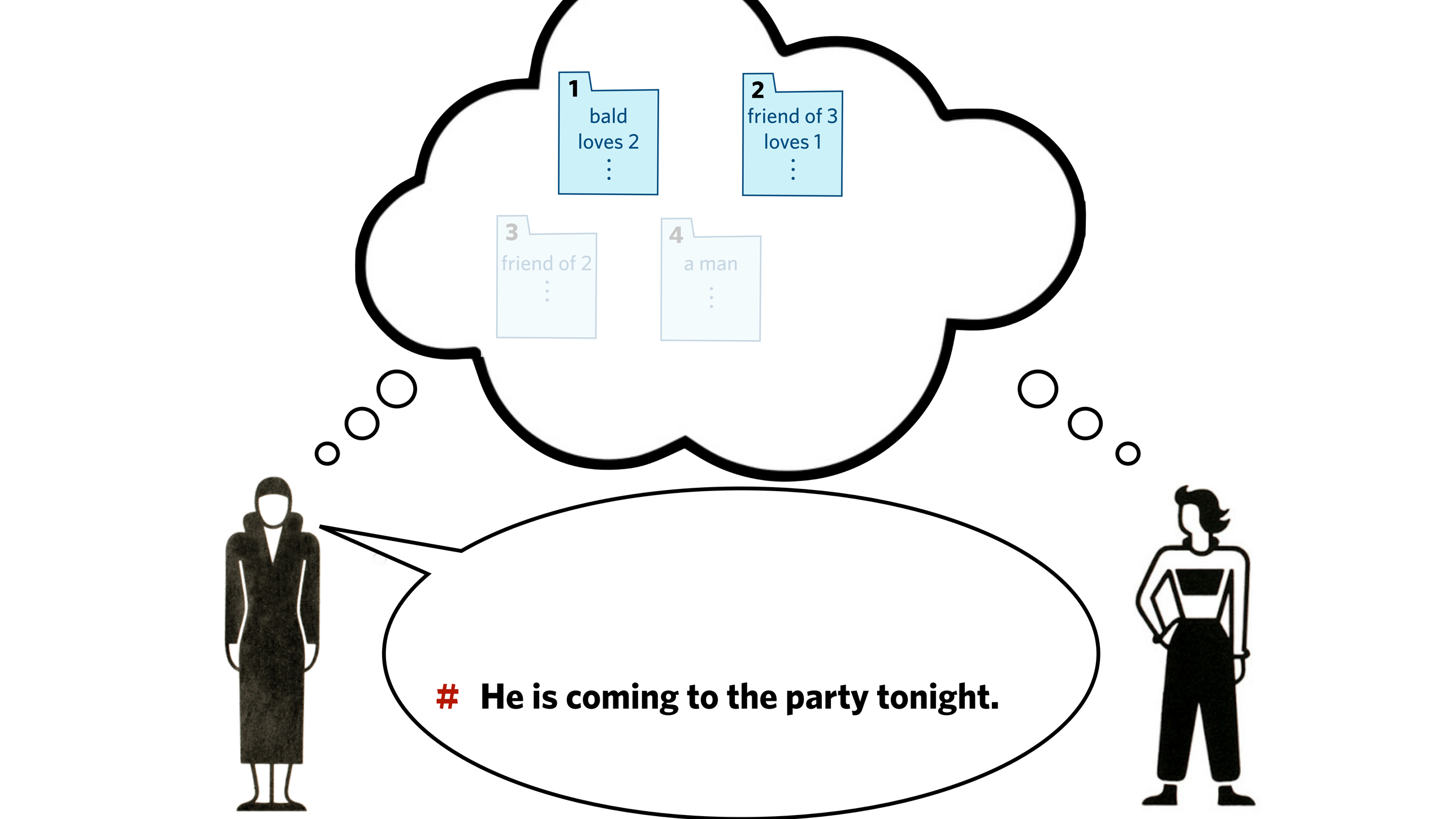


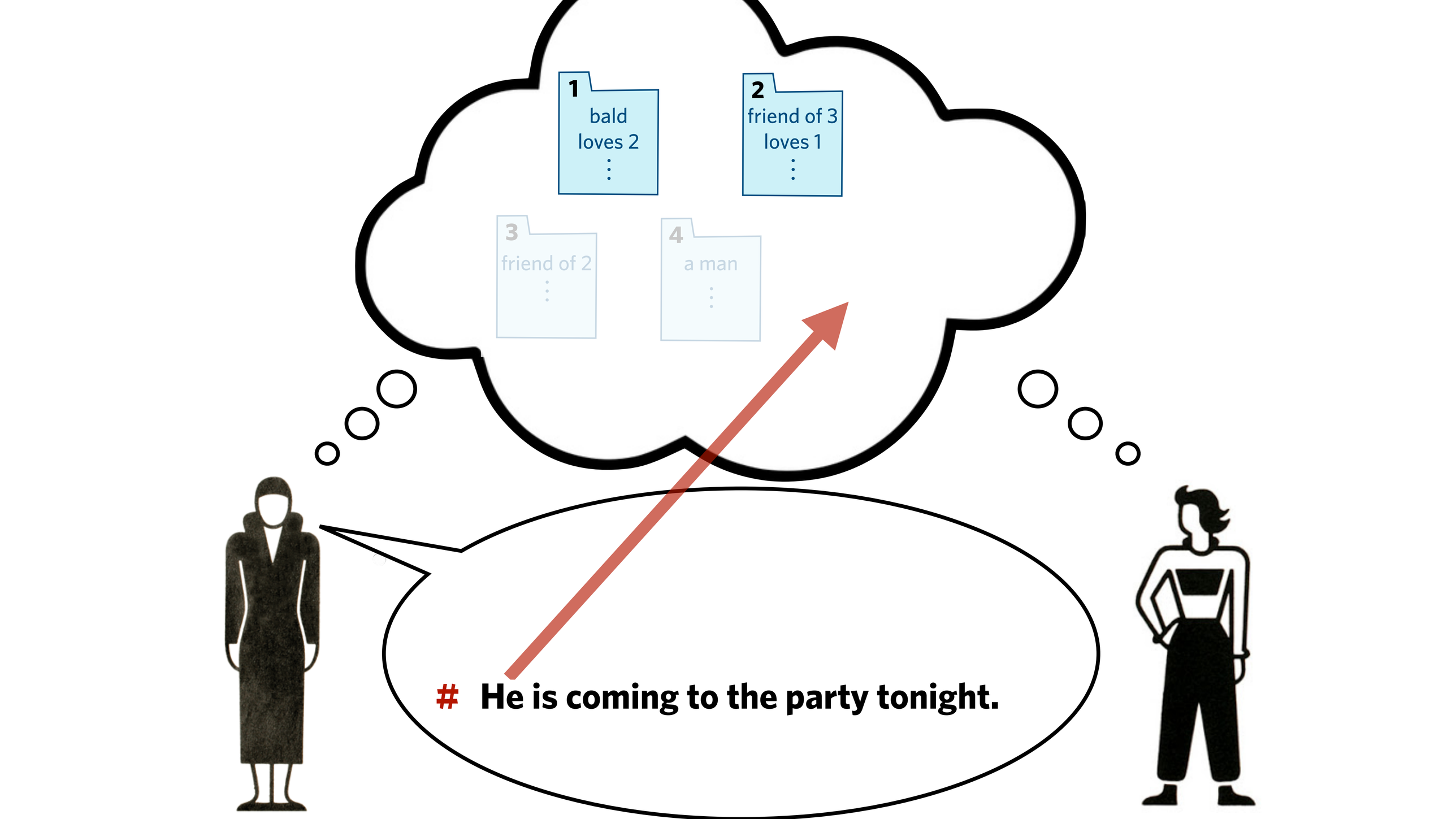
Roberts (2003)

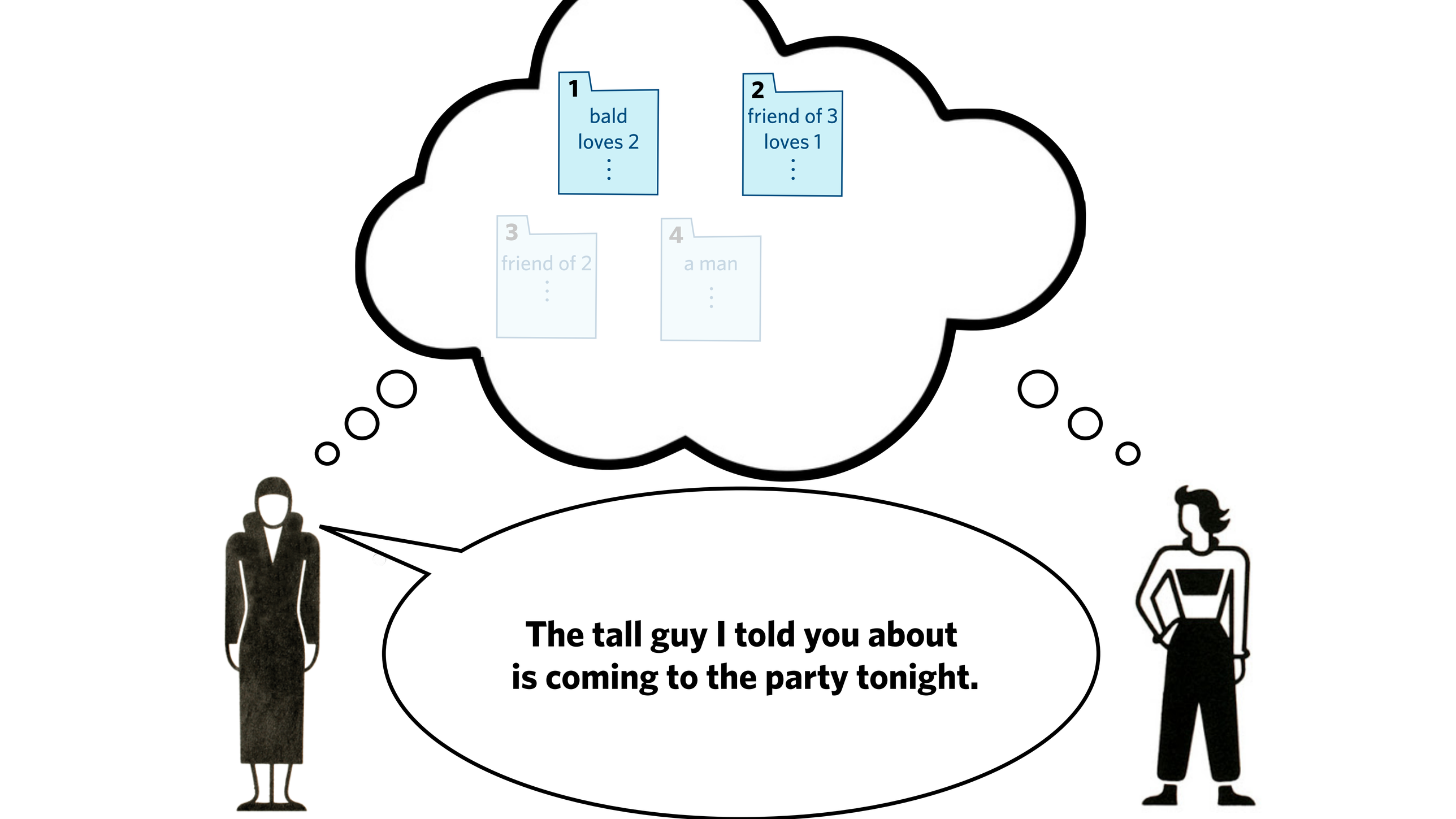

1
bald
loves 2
2
friend of 3
loves 1
3
friend of 2
4
a man
I know a man.

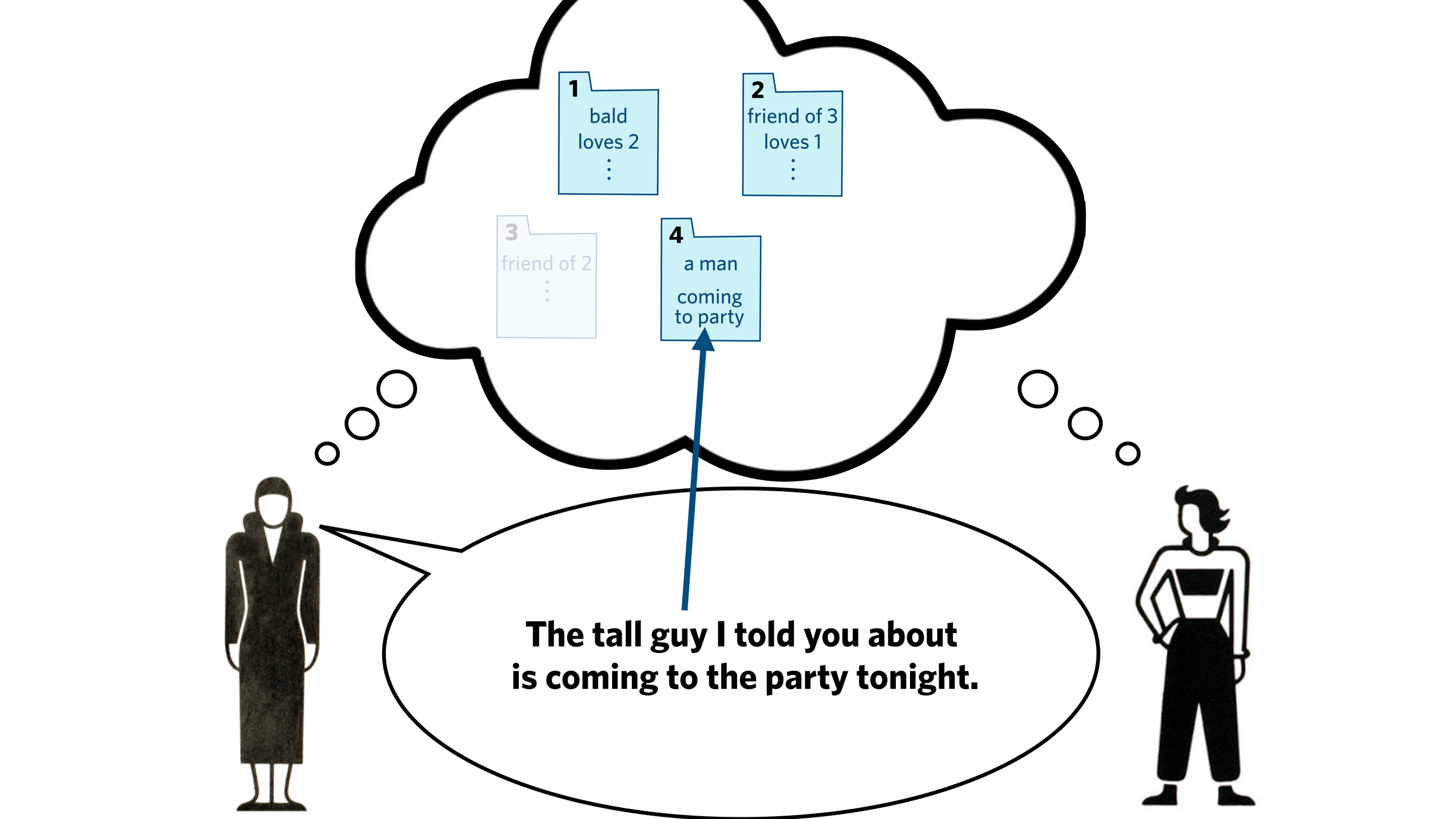

1
bald
loves 2
2
friend of 3
loves 1
3
friend of 2
4
a man
5
a man
known to S
I know a man.


1
bald
loves 2
2
friend of 3
loves 1
3
friend of 2
4
a man
5
a man
known to S
coming to
party
I know a man.
He is coming to the party tonight.


1
bald
loves 2
2
friend of 3
loves 1
3
friend of 2
4
a man
He is coming to the party tonight.


1
bald
loves 2
2
friend of 3
loves 1
3
friend of 2
4
a man
He is coming to the party tonight.


1
bald
loves 2
2
friend of 3
loves 1
3
friend of 2
4
a man
The tall guy I told you about
is coming to the party tonight.


1
bald
loves 2
2
friend of 3
loves 1
3
friend of 2
4
a man
coming
to party
The tall guy I told you about
is coming to the party tonight.

Dynamic Theories: Questions

- How do discourse referents (and their relative salience) relate to our mental representations?
- Are discourse referents models of mental files (/ object files)?
- If so, how do we coordinate them with others?

Dynamic Theories: Questions

- How do discourse referents (and their relative salience) relate to our mental representations?
- Are discourse referents models of mental files (/ object files)?
- If so, how do we coordinate them with others?

Penultimate draft. Please cite published version (when available).

Forthcoming in Oxford Studies in Philosophy of Language, Volume 4. Eds Ernie Lepore and David Sosa

Discourse Referents in a Dynamic Pragmatics

Karen S. Lewis
Department of Philosophy, Barnard College, Columbia University

klewis@barnard.edu

1 Introduction

A formal theory of conversational contexts invariably has to address the question of what elements are represented in a context. Many such theories include *discourse referents* to track anaphoric connections (Kamp (1981), Heim (1982), Kamp & Reyle (1993), Stokhof et al. (1996), Roberts (2003, 2004b), Brasoveanu (2008), a.o.). A discourse referent is like address at which one stores information that hangs together according to the discourse. While discourse referents have been used in accounts of many kinds of anaphora (event, temporal, propositional, etc.), the scope of the present paper is restricted to those that license singular pronominal anaphora or sin-

CONTEXT-SENSITIVITY AND SOCIAL COGNITION

Epistemic Modals

It <u>might</u> rain this week.

How do we coordinate on a body of information to serve as the modal's domain of quantification?

(Is this body itself an object of social cognition?)

CONTEXT-SENSITIVITY AND SOCIAL COGNITION

Definite Reference

Give me this/that.

How do does the speaker choose which expression to use, given that they have a referent?

How do we coordinate on a referent for *that*?

CONTEXT-SENSITIVITY AND SOCIAL COGNITION

Definite Reference

Give me this/that.

How do does the speaker choose which expression to use, given that they have a referent?

How do we coordinate on a referent for *that*?

MUCH MORE TOMORROW!

PREVIEW

TUESDAY
REFERENCE & ATTENTION

WEDNESDAY
EVENT COGNITION

THURSDAY
COMMON GROUND

FRIDAY
FACTIVE MINDREADING