



Cognitive Science 36 (2012) 607–634

Copyright © 2012 Cognitive Science Society, Inc. All rights reserved.

ISSN: 0364-0213 print / 1551-6709 online

DOI: 10.1111/j.1551-6709.2011.01230.x

# Conventional Wisdom: Negotiating Conventions of Reference Enhances Category Learning

John Voiklis,<sup>a</sup> James E. Corter<sup>b</sup>

<sup>a</sup>*Center for Decision Technologies, Stevens Institute of Technology, Hoboken, NJ*

<sup>b</sup>*Department of Human Development, Teachers College, Columbia University*

Received 14 January 2011; received in revised form 27 April 2011; accepted 29 April 2011

---

## Abstract

Collaborators generally coordinate their activities through communication, during which they readily negotiate a shared lexicon for activity-related objects. This social-pragmatic activity both recruits and affects cognitive and social-cognitive processes ranging from selective attention to perspective taking. We ask whether negotiating reference also facilitates category learning or might private verbalization yield comparable facilitation? Participants in three referential conditions learned to classify imaginary creatures according to combinations of functional features—nutritive and destructive—that implicitly defined four categories. Remote partners communicated in the Dialogue condition. In the Monologue condition, participants recorded audio descriptions for their own later use. Controls worked silently. Dialogue yielded better category learning, with wider distribution of attention. Monologue offered no benefits over working silently. We conclude that negotiating reference compels collaborators to find communicable structure in their shared activity; this social-pragmatic constraint accelerates category learning and likely provides much of the benefit recently ascribed to learning labeled categories.

*Keywords:* Reference; Pragmatics; Joint activity; Coordination; Category learning

---

## 1. Introduction

Human beings engage in myriad joint activities: a parent and child jointly build a Lego robot; two families jointly plan a wedding celebration; far-flung scholars and scientists jointly develop a domain of knowledge. Coordination of activity presupposes coordination of intentions, assumptions, and beliefs that drive those activities (Schelling, 1960).

---

Correspondence should be sent to John Voiklis, 211 West 107th Street, #1W, New York, NY 10025. E-mail: voiklis@gmail.com

Adherence to convention—particularly conventions of reference or language games—facilitates both mental and behavioral coordination in repeated activities (Lewis, 1969). For example, cooking is a recurring, often institutionalized, activity with conventional associations between ingredients, tools, and practices. A chef can convey the intention of preparing a North-African stew by commanding “heat up the tagine [a North-African stew pot];” the apprentice can confirm that intention by proposing “I’ll put on a kettle for couscous [a North-African pasta], as well.”

In novel activities, collaborators must coordinate activity as they coordinate how to talk about the activity. Often, these negotiations yield ad hoc conventions of reference or conceptual pacts (respectively, Garrod & Anderson, 1987; Brennan & Clark, 1996). Imagine, for example, the kitchen collaborators repairing their stove’s electric ignitor. Unfamiliar with the conventional names for circuit parts, the apprentice proposes, “that bulb-thingy has burned out,” referring to a glass-cartridge fuse. “I see,” the chef confirms, “the filament has melted.” Through this seemingly trivial exchange, apprentice and chef establish “bulb-thingy” as a referential precedent and, when subsequently repeated, as a conceptual pact. Does this pact entail anything beyond tacit agreement on what to call the unfamiliar fuse? Does it accomplish anything more than facilitate conversation? What conceptual utility accrues from conceptual pacts?

To begin answering these questions and others that follow, we glean insights from multiple related research traditions—psycholinguistics, categorization, social cognition, and language learning in human and machines, among others. We first frame referential communication as negotiated categorization, wherein the processes of negotiating shared labels or expressions for task-relevant objects and/or features of objects (i.e., referential processes) recruit and align the processes of selectively allocating attention, applying prior knowledge, and inferring relevance or meaning (i.e., conceptual processes). Consequently, the externally apparent aspect of conceptual pacts—coordinated referential labels or expressions—bespeaks coordinated internal processes and representations. In terms relevant to the present study, conceptual pacts provide evidence of a negotiated form of category learning. After contrasting the public or explicit reasoning of negotiated category learning against the private or implicit reasoning of individual category learning, we hypothesize that social-pragmatic constraints on negotiating reference will yield better and/or faster category learning than individual processes. We then consider the conceptual effects of using language outside of explicitly communicative or dialogic settings and argue that purely private uses of language likely offer no special advantages over nonverbal individual categorization processes. Finally, we introduce the present study and report on our attempts to disentangle the conceptual effects of communication from the effects of using language, *per se*.

### *1.1. On the relationship between referential and conceptual processes*

Referential processes involve much more than exchanging lexical labels for otherwise self-evident referents. People communicate in order to exert some control over what and how the other perceives and conceives in their shared environment, as well as how the other acts on those perceptions/conceptions (Austin, 1975). For example, when collaborators

negotiate reference to objects in their shared environment, they likewise negotiate categorizations of those objects (Barr & Kronmüller, 2006; Brown, 1958; Cruse, 1977). We do not argue here that lexical choices determine how collaborators conceive objects to which they refer (see Malt, Sloman, & Gennari, 2003; on linguistic vs. non-linguistic categorization). Rather, lexical choices signal social-pragmatic design: lexical choices might point to particular objects or feature (*deixis*), and/or cue common experiences or knowledge (*presupposition*), and/or invite otherwise unsaid meanings or interpretations (*implicature*), and so forth. In all, speakers design referential expressions for interpretation and addressees interpret those expressions assuming design (H. H. Clark & Murphy, 1982; Fussell & Krauss, 1989a, 1989b). Such repeated efforts to design and interpret expressions both recruit and affect cognitive and social-cognitive processes, including selective attention, reasoning, memory, perspective taking, and intention reading (cf. Holtgraves & Kashima, 2008; Pickering & Garrod, 2004; Chiu, Krauss, & Lau, 1998; also Echterhoff, Lang, Krämer, & Higgins, 2009). This joint referential/conceptual process is most evident in developmental (e.g., Tomasello, 2000; E. V. Clark & Amaral, 2010) and robotic (e.g., Steels & Kaplan, 2002) studies of language learning.

One can also see this joint referential/conceptual process at work in the ignitor scenario. The apprentice refers to the glass-cartridge fuse as a “bulb-thingy” to help the chef *differentiate* the glass tube containing a fine metal element from the similarly shaped diodes and resistors (cf. E. V. Clark, 1987). In doing so, the apprentice *presupposes* at least minimal common knowledge of light bulbs (cf. Stalnaker, 2002)—for example, features like *made of glass and metal* and *requires intact filament to function*—and likely expects that knowledge will *direct the chef’s attention* towards circuit parts that possess such features (cf. Brennan, 1995). Similarly, the apprentice can *imply* a broken circuit by pointing to the “burned out” bulb. As one might expect in an idealized example, the chef successfully *infers* that the apprentice has proposed a cause for the malfunctioning ignitor and *confirms the joint construal* of the referring expression by pointing to the melted “filament” (cf. Krauss, 1987; Krauss & Weinheimer, 1966; Wilkes-Gibbs & Clark, 1992). In summary, the apprentice manipulates the chef’s conceptual processes by designing a referring expression that directs the chef’s attention toward features that both differentiate the target from other possible referents and allow the chef to infer the referent’s significance to the activity. The chef’s confirmatory remark completes the negotiated categorization of the fuse and its significance to the joint activity.

### 1.2. On the relationship between referential and conceptual structure

What conceptual utility accrues from conceptual pacts? Conceptual pacts appear to stabilize the conceptual effects of negotiating reference. When interlocutors adhere to a shared pattern of referential labels and expressions (i.e., *referential structure*), they signal mutual adherence to a shared pattern of selective attention, construals, and inferences (i.e., *conceptual structure*). By the principle of least effort (Clark, 1996), adhering to conceptual pacts helps interlocutors more easily direct joint attention, confirm joint construal, and execute joint intentions. More important for the present study, these conceptual effects persist: Following conversation, conceptual pacts influence how each collaborator sorts the object

referents of the pact (Markman & Makin, 1998) and how each collaborator later judges the similarity and/or typicality of objects to categories named in accordance with the pact (Malt & Sloman, 2004). In other words, negotiating reference imposes structure on a previously undifferentiated (or less differentiated) novel task environment (e.g., circuitry, in the case of the kitchen collaborators). The shareability hypothesis (Freyd, 1983) posits that category structure emerges as people share category knowledge and to enable the sharing of such knowledge; that is, people represent (at least, socially derived) knowledge in a shareable form. These persistent conceptual effects suggest that negotiating reference functions concurrently as a negotiated form of category learning. Therefore, we ask whether negotiating reference with a collaborator actually *facilitates* category learning: Does the search for communicable structure in a shared or public activity yield better and/or faster category learning than one might expect from individuals engaged in a similar but *private* activity?

Research on the conceptual effects of communication suggests that active participants in conversation exploit different reasoning processes and strategies from those outside of conversation (see review by Brennan, Galati, & Kuhlen, 2010). The conceptual structure to which conceptual pacts refer often remains opaque to those overhearing but not engaged in a conversation (Schober & H. H. Clark, 1989) or to those stuck in a one-sided conversation with no opportunity to negotiate meaning, such as interviewees (Schober & Conrad, 1997). Mutual understanding requires the social-pragmatic cues afforded by dialogue (cf. Brown-Schmidt, 2009). In fact, people who exhibit impairments in learning (e.g., amnesia; Duff, Hengst, Tranel, & Cohen, 2005) and using (e.g., aphasia; Hengst, 2003) semantic knowledge during private activity can still learn and share new semantic knowledge during dialogue. Conversely, people who exhibit social-pragmatic impairments (e.g., autism) have little trouble finding structure in their private activities (e.g., J. Brown, Aczel, Jiménez, Barry, & Plaisted, 2010), but this facility appears to reflect statistical (implicit) rather than categorical (explicit) reasoning (Gastgeb, Strauss & Minshew, 2006; Soulières, Mottron, Saumier, & Larochelle, 2007). As observable in connectionist models (see critique by A. Clark & Karmiloff-Smith, 1993), implicit reasoning yields structures that often defy verbalization (required for communication), much less analysis (required for design and interpretation). We do not here equate private activity with autism, but, absent the knowledge-sharing imperative, typical individuals likely organize their private activities using more or less the same implicit reasoning processes as autistic individuals. On the other hand, dialogue entails on-the-fly transactive analysis of activity-related and social-pragmatic information that outpaces similar individual processes (see Mercier & Sperber, 2011; Sperber & Mercier, in press, on reasoning as a social skill). We do not doubt that typical individuals can discern the same structures in their private activities as collaborators discern in their public activities (as argued by Malt et al., 2003), but individual processes may take more time and yield less differentiated structure than dialogue.

### 1.3. Some conceptual consequences of labeling categories

That said, similar conceptual facilitation might derive from simply verbalizing reasoning processes without explicit communication (cf. Vygotsky, 1962/1986). For example, both

children (e.g., Waxman & Markow, 1995) and adults (e.g., Lupyan, Rakison, & McClelland, 2007) learn to differentiate labeled categories more quickly than unlabeled categories. Moreover, labeling categories appears to maximize attention to name-relevant features, directing visual search (Lupyan, 2008a) and recognition (Lupyan, 2008b) processes among adults and helping children infer unknown object features (Gelman & Markman, 1986) and generalize labels to novel objects (Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002). Relational labels, in particular, focus attention on how perceptual features relate to one another (Gentner, 2003) or to functional features (Mueller Gathercole, Cramer, Somerville, & Jansen op de Haar, 1995) and help children better remember such relationships (Dessalegn & Landau, 2008). These and other studies suggest that labeling itself, whether used in public or private contexts, enhances conceptual representation and reasoning (Gentner & Goldin-Meadow, 2003).

Obviously, public and private uses of language have much in common: it seems difficult, *almost* nonsensical, to study the conceptual effects of referential communication while ignoring the effects of referential labels, whether the experimenter or the participants devise the labels. It seems equally difficult, perhaps impossible, to study the conceptual effects of labeling in the absence of actual or implied communication; the lack of an explicitly defined communication paradigm between experimental subjects does not preclude implicit communication between the labeler (the speaker) and the audience for that label (the addressee). In some cases, the investigator might serve as speaker and experimental subjects as addressees; in other cases, the roles may reverse. For example, Lupyan et al. (2007) provided subjects with two nonsense labels for two kinds of imaginary aliens. Labeling the aliens likely functioned as implicit communication, prompting subjects to look for what difference the two contrasting labels were intended to communicate. Similarly, Lupyan (2008b) instructed subjects to label objects using basic-level terms—for example, “chair” as opposed to “recliner.” Basic-level terms convey pragmatically useful information about the referent (in terms of the Maxim of Quantity; Grice, 1975); instructions to use such terms communicate the need to focus on that information. In these and other studies, subjects do not rely on a private lexicon, rather implicit communication establishes an implicit conceptual pact between subject and investigator. Purely private referential processes may fail to produce results comparable to these implicit pacts (cf. Wittgenstein, 1958/2001).

#### *1.4. Talking to oneself does not require pragmatics*

Prior comparisons of public and private reference appear, at first, to support an opposing view: People seem better able to recognize the referent of their own expressions than the referent of expressions written for some generic other (Fussell & Krauss, 1989a, 1989b; Krauss, Vivekananthan, & Weinheimer, 1968). We contend, though, that private referential expressions might convey enough conceptual information for recognition, but not necessarily enough for categorization; recognition falls far short of categorization and likely entails different cognitive processes (cf. Knowlton & Squire, 1993). Moreover, writing messages for a generic other or one’s future self certainly qualifies

as communication, but falls far short of dialogue. As noted earlier, dialogue involves collaborative processes wherein speakers design and redesign referential expressions for particular addressees; addressees, in turn, provide concurrent feedback on whether and how they interpret those expressions. Both speaker and addressee derive conceptual utility from these processes of design and interpretation. Private reference or monologue functions as an one-sided conversation and, consequently, exhibits minimal interpretable design (Fussell & Krauss, 1989a, 1989b; Krauss et al., 1968). Consequently, we do not expect monologue to yield conceptual utility comparable to dialogue.

### 1.5. Getting past the nonsense: The present study

With the present study, we attempt to start disentangling the conceptual effects of *communication* (the social and pragmatic aspects of language use) from the effects of using language per se (the private use of labels and other referring expressions). To do this, we compare the conceptual consequences of negotiating shared reference to the consequences of inventing a private lexicon, and ask if the dialogue better facilitates category learning and category use. We expect that negotiating shared reference during dialogue (more so than either monologue or silent learning) will compel interlocutors to search for shareable structures upon which to design and interpret referential expressions. This social-pragmatic effort will [H2] widen the distribution of attention to features and highlight how features relate to one another and to the latent properties of referents (e.g., their functional significance). Consequently, we expect that negotiating shared reference during dialogue will [H1] enhance category learning; inventing a private lexicon during monologue will offer no benefits over silent, individual learning.

To test these hypotheses, we designed a function-prediction task to encourage the indirect learning of implicitly defined categories (see Markman & Ross, 2003; on indirect category learning). We implemented the task as a computer game, in which players predict whether various alien creatures provide food ( $\pm$ *nutritive*) and/or present a threat ( $\pm$ *destructive*). The conjunction of these orthogonal functions implicitly defines four functionally and perceptually distinct categories. Otherwise, categories lack labels and corrective feedback addresses each function independently. Function prediction entails two distinct roles: the *spotter*, who views the creature; and the *beamer*, who performs function-appropriate actions. These roles permit play under differing referential conditions. Prediction requires *dialogue* when collaborative players alternate between spotter and beamer. When a single player alternates between roles, prediction requires *monologue* (recording audio messages for later listening<sup>1</sup>). Talking is unnecessary when a single player performs both roles at once (*control*). For this study, we limited talk to descriptions of observable features; explicit reference to actions or functions was prohibited. These referential conditions and constraints allowed us to estimate the simple effects on category learning of lexical invention during monologue, and the effects of negotiating reference during dialogue.

## 2. Method

### 2.1. Participants

Forty-six male and 50 female students (median age 23) from throughout the Columbia University community participated in this study for a cash payment and the chance to win a digital music player (awarded to top performers in each condition). Participants were recruited using flyers posted across the campus. All participants were native speakers of English, with an average of 4-year post-secondary schooling during which they devoted approximately 2 h per week to computer games.

### 2.2. Design

Participants were randomly assigned to one of three referential conditions. Remote partners communicated in the Dialogue condition ( $N = 42$ ). In the Monologue condition ( $N = 20$ ), participants recorded audio descriptions for their own later use. Controls worked silently ( $N = 34$ ). To isolate social-communicative effects on category learning from the effects of private speech, we analyzed these data using two orthogonal contrasts of these conditions. *Communication* contrasted dialogic participants against individuals (both Monologic participants and Controls). *Speaking* contrasted Monologic participants against Controls.

### 2.3. Materials

Participants repeatedly classified 16 alien creatures. Table 1 depicts, in binary notation, the structural aspects of one specific set of creatures. Creatures varied on five binary-valued observable features: Tentacles, Fins, Heart, Eyes, and Color. For example, creatures possessed either pointed or rounded fins. In addition, for naturalistic noise, the Size of each creature varied randomly from 90% to 110% of its standard measurement.

Participants classified creatures on two functionally significant features (criterion variables): whether a creature produces jelly usable as food and fuel ( $\pm$ *nutritive*) and whether it might damage life support systems ( $\pm$ *destructive*). The joint prediction of these orthogonal functions implicitly defined four unlabeled categories of creatures: nutritive and destructive (ND), nutritive only (Nx), destructive only (xD), or no functional significance (xx).

In the particular category structure depicted by Table 1, the types of Tentacles, Fins, and Heart define a family resemblance structure that predicts whether the creature is nutritive. A value of “1” on at least two of these three features means the creature is nutritive, otherwise not. By contrast, Eye-type can serve as a simple rule to predict whether the creature is destructive (see Minda & Ross, 2004, on category structures that combine a simple rule with a family resemblance structure). As a counterbalancing measure, we designed three additional category structures, assigning three permutations of Tentacles, Fins, Heart, and Eyes to the predictive dimensions. We also counterbalanced which structure type—family

Table 1

The category structure used in the experiment, which illustrates one of four possible assignments of diagnostic surface features (tentacles, fins, heart, eyes) to category structure

Exemplar	Observable Dimensions						Functional Dimensions		Category
	Diagnostic				Non-Diagnostic		Nutr.	Dest.	
	Tent.	Fins	Heart	Eyes	Color	Size			
ND.1	1	1	1	1	1	Random	1	1	ND
ND.2	1	1	0	1	1	Random	1	1	ND
ND.3	1	0	1	1	0	Random	1	1	ND
ND.4	0	1	1	1	0	Random	1	1	ND
Nx.1	1	1	1	0	0	Random	1	0	Nx
Nx.2	1	1	0	0	0	Random	1	0	Nx
Nx.3	1	0	1	0	1	Random	1	0	Nx
Nx.4	0	1	1	0	1	Random	1	0	Nx
xD.1	1	0	0	1	1	Random	0	1	xD
xD.2	0	1	0	1	1	Random	0	1	xD
xD.3	0	0	1	1	0	Random	0	1	xD
xD.4	0	0	0	1	0	Random	0	1	xD
xx.1	1	0	0	0	0	Random	0	0	xx
xx.2	0	1	0	0	0	Random	0	0	xx
xx.3	0	0	1	0	1	Random	0	0	xx
xx.4	0	0	0	0	1	Random	0	0	xx

resemblance or simple rule—predicted the nutritive or destructive function. Color and Size were never used as predictive features and did not correlate with any other feature.

## 2.4. Procedure and Tasks

### 2.4.1. Pretest

Before training, all participants individually performed a *Free Sort* of the 16 creatures without any knowledge of functions or the prediction task to follow (Fig. 1B shows the free sort interface). At the start of the task, participants encountered all 16 creatures randomly arranged in a single container. While in a container, creatures were rendered at 15% of their normal size to fit on the computer screen; dragging a creature onto the desktop allowed participants to view it at full size. Participants were encouraged to inspect each creature at full size and familiarize themselves with all of its observable features before deciding on how to sort it. To sort creatures, participants created a number of graphical containers into which they dragged the creatures they believed belonged together. Upon dropping a creature into a newly created sorting container, an explanation field appeared in the container into which participants were instructed to provide a short explanation of why the creatures in that container belonged together. Participants could compose or edit that explanation at any time during the sorting process. Participants were free to create as many or as few categories (from 1 to 16) as they deemed necessary. The pretest free sort provided data on the *number of features mentioned* by participants when explaining creature groupings.



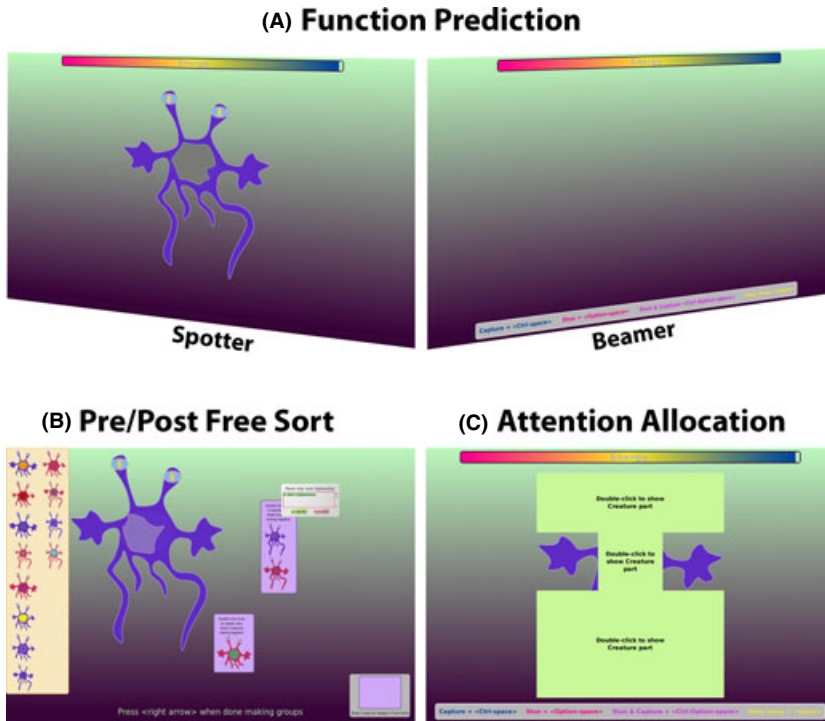


Fig. 1. (A) The interface for the Function Prediction task: the *spotter* (left) could see the creature, and the *beamer* (right) performed the predictive actions (capturing and/or stunning creatures). The single-player version (Monologue & Control) integrated the spotter and beamer interfaces into a single interface. (B) Pre/Post-training Free Sort interface: participants sorted creatures into as many graphical bins as they deemed necessary and explained why co-occurring creatures belonged together. (C) The interface for the Attention Allocation posttest: resembles single-player Function Prediction interface, except that participants uncover (remove graphical blinds) a creature feature-by-feature before predicting its function.

#### 2.4.2. Training

Participants imagined that they were preparing for a mission to a planet populated by creatures that might or might not produce food and fuel and might or might not attack. The “game” would train them to predict creature functions. Training proceeded over 320 trials: 10 blocks of 32 trials (i.e., two randomly ordered presentations of the stimuli per block).

*Function Prediction* entailed two roles: the *spotter*, who viewed the creature; and the *beamer*, who predicted functions. Dialogic partners alternated (on every trial) between spotter and beamer<sup>2</sup> as they collaborated through networked workstations divided by a 5' × 5' barrier; they could hear but not see one another. When playing as beamer, dialogic participants could not see the creatures on which they acted; instead, their spotters described the creature using as many observable features as they deemed necessary for prediction, without referring explicitly to actions or functions. Beamers could seek clarification. Training lasted for 320 trials. Each dialogic partner alternately described creatures on 160 trials and predicted functions on the other 160 trials.

Monologic participants alternated (on every block of 16 trials) between spotter and beamer as they collaborated with themselves. On half the trials, they recorded audio descriptions like those of dialogic spotters, then immediately predicted the functions. On other trials, they predicted functions after hearing recent descriptions (recorded during the previous 16-trial block) of creatures they could not see. In this way, Monologic participants experienced each of the prediction roles (spotter and beamer) and each of the conversational roles (speaker and addressee). Overall, Monologic participants predicted functions on all 320 trials; they described creatures on every other block of 16 trials and listened to those descriptions on during the intervening blocks of 16 trials.

Controls performed both the spotter and beamer roles at once. They predicted functions on all 320 trials without talking.

On each training trial, 1 of 16 creatures appeared at the center of the spotter's video display, where it remained until the beamer acted or 20 s (the maximum trial length) elapsed. The beamer executed function-related actions—*stunning* destructive creatures and/or *capturing* nutritive creatures—using two keystroke combinations on a standard computer keyboard (Fig. 1A shows the combined beamer/spotter interface).

In response to key combinations, a positive or negative tone signaled overall accuracy (i.e., correct or incorrect predictions of both functions combined). Then, corrective feedback reinforced (partially) correct predictions, while correcting mistakes. Specifically, a synthesized voice described the function-related consequences of the chosen combination. For example, after correctly capturing an Nx creature, players heard “jelly extracted”; alternatively, players who mistakenly stunned and captured an Nx creature heard “stun beam wasted, some jelly extracted.” Finally, a graphical *energy meter* further reinforced the descriptive feedback by increasing or decreasing its length. To increase motivation in this potentially difficult training task, we offered a prize (a digital music player) to the best performing participant(s).

The Function-prediction Training task provided data on *correct and incorrect predictions of individual functions and function conjunctions* (i.e., categories), as well as *the number and order of features* to which participants referred in the Monologue and Dialogue conditions.

#### 2.4.3. Posttests

After training, participants worked alone on two posttests. First, participants performed an additional *Free Sort* of creatures and provided short explanations for each grouping of creatures. They were instructed to use the knowledge gleaned from the training task; otherwise the posttest and pretest sorts were identical. The Posttest Free Sort provided data on *which creatures were grouped together* by participants, as well as the *number and kind* (observable, functional, or behavioral) *of features mentioned* by participants when explaining creature groupings.

Finally, participants from all conditions worked individually on an *Attention Allocation* task that we intended to elicit data on selective attention to diagnostic features during predictions. This task resembles the single-player function-prediction training task in all aspects except that each creature appears with its various diagnostic features hidden behind

graphical blinds (Fig. 1C shows the Attention Allocation interface). Participants physically *uncovered* (by mouse-clicking the blind) as many features as they desired before selecting an action. Selective attention data entailed both the *number* and *order of features uncovered* by participants.

### 3. Results

We have organized the results around our two hypotheses: [H1] dialogue enhances category learning; and [H2] dialogue broadens the allocation of attention to features and yields greater attention to how features relate to functions. In support of Hypothesis 1, we report analyses of category and function prediction data, the structure of sort data, and the explanations participants offer to justify their sort clusters. In support of Hypothesis 2, we report analyses of selective attention to features and relationships between features and functions. We then link these patterns of selective attention to patterns of reference. We conclude by presenting some additional analyses of an alternative hypothesis: Specifically we ask whether information pooling between interlocutors might explain the wider distribution of Dialogic attention and better category learning.

#### 3.1. H1. Dialogue enhances category learning

We assessed category learning with data from the Function-prediction Training task and the Posttest Free Sort task. The prediction data allow us to compare category-learning *efficiency*—that is, how quickly participants appear to learn the experimenter-defined category structure. The sort data allow us to compare category-learning *efficacy*—that is, how well participants ultimately learn that structure.

##### 3.1.1. Dialogue yields more efficient category learning

The training task provided data on correct and incorrect predictions of individual functions ( $\pm$ *nutritive* and  $\pm$ *destructive*). One can infer learning of the four categories (implicitly defined by the conjunction of the two functions) by measuring when participants correctly predict the conjunction of the two functions of the stimulus creature. We coded category predictions as “1” when both function predictions were correct and “0” otherwise. We derive our measure of “category-prediction accuracy” by computing the proportion of correct category predictions within each of the 10 blocks.<sup>3</sup>

*3.1.1.1. Category-prediction accuracy:* The advantage for Dialogic dyads<sup>4</sup> is evident in Fig. 2, which compares category-prediction accuracy by referential context across the 320 training trials (10 blocks of 32 trials). Dialogic dyads predicted functions with increasingly greater accuracy during training than did individuals. Monologue participants and Controls both exhibited low (and similar) levels of accuracy.

We corroborated these differences using a repeated measures ANOVA on (arcsine-transformed) “category-prediction accuracy,” using the orthogonal between-groups contrasts

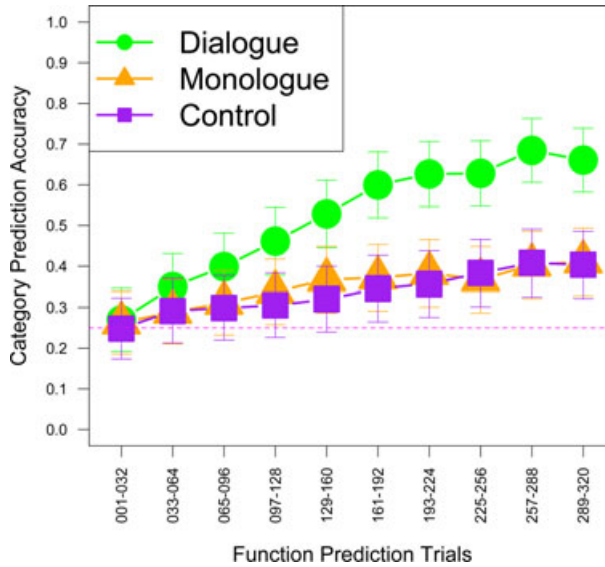


Fig. 2. A comparison of the category prediction accuracy of participants in the three referential conditions across 320 training trials. Error bars indicate 95% confidence interval. Dotted horizontal line indicates chance.

defined in the Design: *Communication* (comparison of Dialogue vs. Monologue and Control) and *Speaking* (comparison of Monologue vs. Control), with Block as a within-subjects factor. The interaction of Communication and Block was significant,  $F(9, 648) = 14.6594, p < .001, \eta_p^2 = .17$ , as were the main effects of Communication,  $F(1, 72) = 33.86, p < .001, \eta_p^2 = .32$ , and Block,  $F(9, 648) = 38.87, p < .001, \eta_p^2 = .35$ . Neither the main effect of Speaking nor the interaction of Speaking and Block reached significance—both  $\eta_p^2 < .01$ . Across training trials, accuracy increased more quickly in the Dialogue condition than in the individual learning conditions.

3.1.1.2. *Function-prediction accuracy:* As described in Materials, participants could predict one of the two functions based on its perfect correlation to one observable feature (simple rule); they could probabilistically predict the other function based on three of the remaining observable (family resemblance) features. It would also seem much easier to establish reference to the single simple-rule feature, compared to the complex of family-resemblance features. If so, then we might expect an early performance advantage for Dialogic dyads in learning to predict the function related to the simple rule structure. We examine this possibility by measuring “function-prediction accuracy”: We coded individual function predictions as “1” when correct and “0” when incorrect, then computed the proportion of correct function predictions within each of the 10 blocks of 32 training trials.

Participants in all referential conditions predicted the function related to the simple rule structure (see Fig. 3A) with greater accuracy than they predicted the function related to the family resemblance structure (see Fig. 3B). Nevertheless, Dialogic dyads exhibited an earlier and greater advantage in predicting the simple-rule function than did Monologic

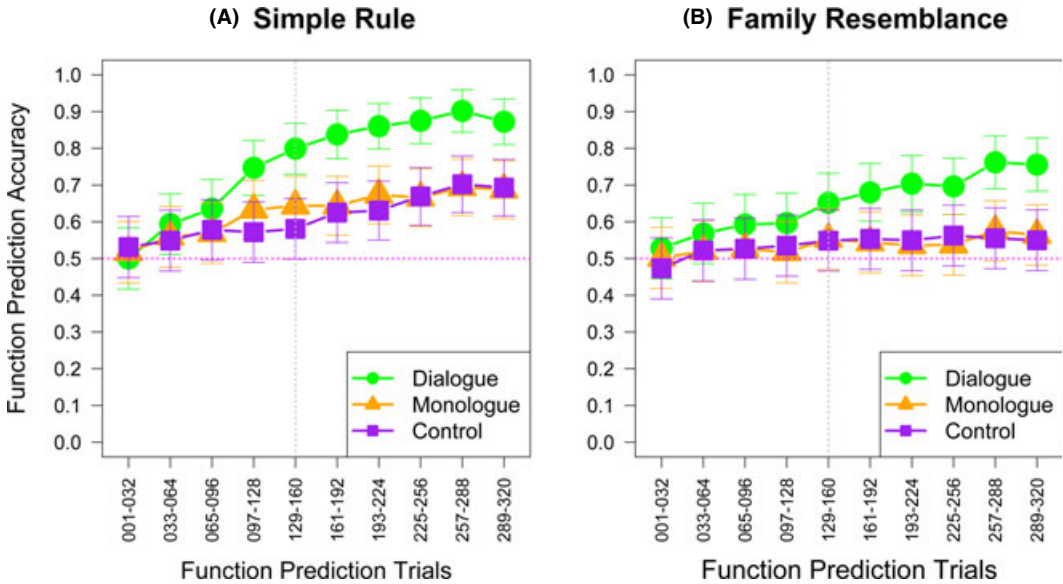


Fig. 3. (A) Comparison of the three referential conditions on prediction accuracy for the function corresponding to the simple rule (perfect correlation between one feature and one function) across 320 training trials. (B) Comparison of the three referential conditions on prediction accuracy for the function corresponding to the family resemblance structure (probabilistic relationship between three features and second function) across 320 training trials. Error bars indicate 95% confidence interval. Dotted horizontal line indicates chance. Dotted vertical line indicates midpoint of the training task (trial 160).

participants and Controls. The accuracy with which Dialogic dyads predicted the family resemblance function increased mainly during the latter half of training (Trials 161–320). Monologic participants and Controls predicted simple-rule function with slowly increasing accuracy across all training trials (Trials 1–320). Their accuracy in predicting the family-resemblance function hovered just above random response (0.50) throughout training.

A repeated measures ANOVA confirmed these differences (using arcsine-transformed “function-prediction accuracy” values). The effects of Communication were significant in a three-way interaction with Structure-Type (family resemblance vs. simple rule) and Block,  $F(9, 648) = 3.79, p < .001, \eta_p^2 = .05$ ; the advantage for the Dialogue condition grew more rapidly for the simple-rule function than for the family-resemblance function. Communication also yielded a significant two-way interaction with Block,  $F(9, 648) = 14.20, p < .001, \eta_p^2 = .16$ ; the advantage for Dialogue increased over time. Nevertheless, the interaction of Communication and Structure-Type failed to reach significance,  $F(1, 72) = 2.99, p = .09, \eta_p^2 = .04$ . As one can infer from the category prediction results, the main effect of Communication was highly significant,  $F(1, 72) = 27.57, p < .001, \eta_p^2 = .28$ ; dialogic participants predicted functions with greater overall accuracy than individuals.

In addition, Structure-Type (family resemblance vs. simple rule) had significant effects in interaction with Block,  $F(9, 648) = 13.37, p < .001, \eta_p^2 = .16$ , and as a main effect,  $F(1, 72) = 47.49, p < .001, \eta_p^2 = .40$ . The main effect of Block was also significant,  $F(9,$

648) = 44.83,  $p < .001$ ,  $\eta_p^2 = .38$ . Function prediction accuracy generally increased over time, but accuracy increased more quickly for simple-rule predictions.

The effects of Speaking failed to reach significance, whether as a main effect ( $\eta_p^2 < .01$ ), in a three-way interaction with Structure-Type and Block ( $\eta_p^2 = .01$ ), or in interactions with either Structure-Type or Block (both  $\eta_p^2 < .01$ ). Whether describing creatures to themselves or working silently, individual learners did not exhibit significantly different levels of accuracy.

These analyses show that negotiating reference enhances the learning of both simple-rule and family-resemblance structures. Nevertheless, this dialogic advantage appears earlier for simple-rule predictions than for family-resemblance predictions.

### 3.1.2. Dialogue yields better category learning

Some might argue that this asymmetry in learning to predict the simple-rule function versus the family-resemblance structure would suggest that, rather than learning categories, dialogic dyads learned to predict each function separately. The asymmetry might have “blocked” participants from integrating the two function-related substructures into a coherent category structure. One needs to look at how participants sorted creatures after training for more direct evidence of whether participants learned categories.

The Posttest Free Sort data provide that direct evidence of category learning. We devised a measure of correspondence, category fidelity, between the sort clusters created by participants and the four “true” (experimenter-designed) categories. First, we converted participant-created clusters of creatures and the four true categories into binary co-occurrence matrices. If a “row” creature and a “column” creature were sorted into the same group, we entered “1” in the intersecting cell and “0” otherwise. We then rearranged the lower triangle of each matrix as a vector. Finally, we used the normalized mutual information (Krippendorff, 1986) between each of the various co-occurrence vectors to measure similarity between clusters.

Additionally, participants provided short explanations of why creatures in the container belonged together. Participants cited individual observable features, individual functions, and categories (conjunctions of functions or predictive behaviors, such as *capture & stun*, etc.). We coded each citation of a functional and/or behavioral category as “1” and “0” otherwise. We used two coders (one blind to the purpose of our research) to code the feature sequences. Agreement was high (Krippendorff’s  $\alpha = .90$ ), and disagreements were resolved through discussion. The average proportion of category citations across sort clusters measured “category avowal”: the extent to which participants sorted on categories as their declared organizing principle. Together, the somewhat redundant measures of category fidelity and category avowal provide mutually reinforcing evidence of category learning.

As apparent in Fig. 4A, participants trained under the Dialogic conditions sorted creatures into clusters that better resembled the “true” category clusters (greater “category fidelity”) than did the clusters produced by individual learners,  $M(\text{dialogue}) = 0.38$  ( $SD = 0.43$ ) versus  $M(\text{monologue}) = 0.06$  ( $SD = 0.08$ ) and  $M(\text{control}) = 0.09$  ( $SD = 0.14$ ). Moreover, in explaining their sorts (Fig. 4B), Dialogic participants cited functional and/or behavioral categories with greater likelihood (greater “category avowal”) than did individuals,  $48 \pm 16\%$

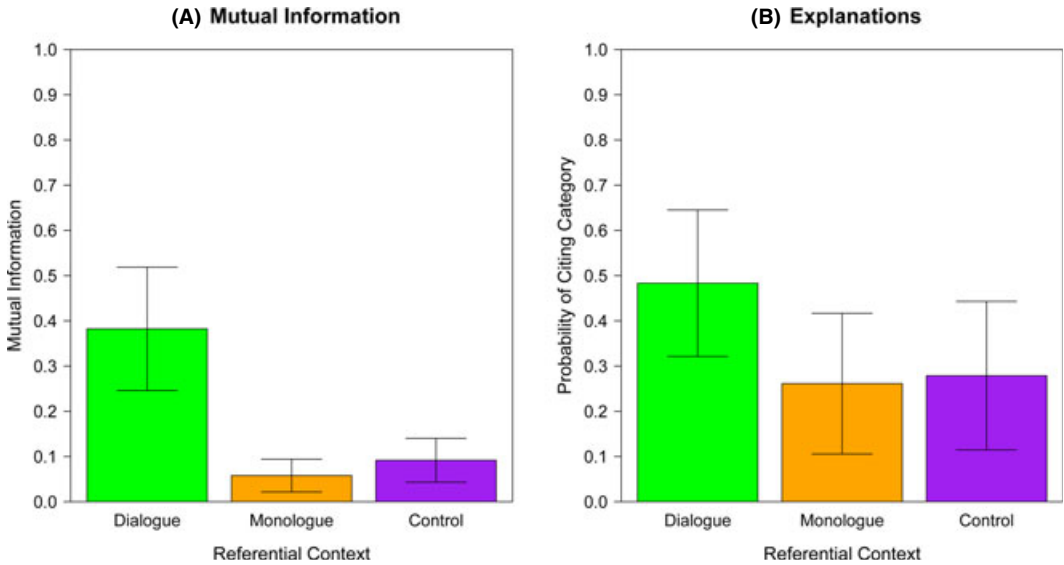


Fig. 4. (A) Comparison of the similarity (category fidelity) of the post-training sort clusters produced by participants in the three referential conditions to the “true” category clusters. (B) Comparison of the probability of citing a category (category avowal: whether functionally defined or behaviorally defined) when participants in the three referential conditions explained their post-training sort clusters. Error bars indicate 95% confidence interval.

(*dialogue*) versus  $26 \pm 16\%$  (*monologue*) and  $28 \pm 16\%$  (*control*). Category fidelity correlated significantly with category avowal,  $r = .52$ ,  $t(94) = 5.83$ ,  $p < .001$ .

A multivariate analysis of variance (MANOVA) on “category fidelity” and (arcsine-transformed) “category avowal” in the sort explanations confirmed these advantages for the participants in the Dialogue condition. Communication produced both a significant multivariate effect,  $F(1, 93) = 11.89$ ,  $p < .001$ ,  $\eta_p^2 = .21$ , and significant univariate effects on category fidelity,  $F(1, 93) = 24.05$ ,  $p < .001$ ,  $\eta_p^2 = .21$ , and on category avowal,  $F(1, 93) = 5.79$ ,  $p = .02$ ,  $\eta_p^2 = .06$ . Neither the multivariate nor univariate effects of Speaking on category fidelity or on category avowal reached significance, with all  $\eta_p^2 < .01$ . In all, dialogic participants, more than individuals, explicitly sorted in accordance with the “true” category clusters.

### 3.1.3. Summary of category learning results

Overall, the category learning results demonstrate that Dialogue yielded better and more efficient learning than training individually. Dialogic participants predicted both individual functions and function conjunctions with greater accuracy than individuals, and they more often appear to have discovered the function-defined category structure.

## 3.2. H2. Dialogue widens selective attention across features and feature relations

Early in this article, we argued that interlocutors manipulate one another’s conceptual processes by designing referring expression that direct the addressee’s attention toward

features that both differentiate the target from other possible referents and allow the addressee to infer the referent's significance to the activity. More simply, we hypothesize that negotiating reference influences patterns of selective attention. This influence points to one mechanism (among other potential mechanisms, see Garrod & Pickering, 2009, for review) through which negotiating reference can facilitate category learning.

We assessed patterns of selective attention with data from the Attention Allocation Posttest and from the referential expressions recorded during Dialogic and Monologic training. These include data on the number and order of overall and family resemblance features that participants either uncovered and/or mentioned when predicting functions. (Please note that "uncover" denotes the physical behavior of removing graphical blinds to expose a hidden creature feature-by-feature during the Attention Allocation task.)

We use the data on uncovered features to assess directly whether the different referential conditions yield different patterns of selective attention. We use the data on mentioned features to (inferentially) link these attentional/conceptual differences to referential processes. These related behaviors of mentioning and uncovering features provide converging evidence of how referential conditions affect selective attention.

### 3.2.1. Dialogue widens the distribution of attention across features

To predict *both* functions with reliable accuracy, participants needed to uncover a minimum of three features: one for the function predictable by a simple unidimensional rule, and *at least* two of the three features that probabilistically predict the other function (which yields an average accuracy of 81.25%). Further, uncovering at least three features might imply that participants have learned to distinguish the four categories implicitly defined by the two functions. So, we computed the "likelihood of *uncovering* three or more features" (i.e., "1" for participants who uncovered at least two family resemblance features and the one simple-rule feature and "0" otherwise). Likewise, using the referring expressions recorded during the last block of training, we computed the "likelihood of *mentioning* three or more features." We use logistic regression to model the binary measures of uncovering/mentioning three or more features.

During the individual posttest,  $67 \pm 7\%$  of Dialogue participants met or exceeded that three-feature minimum, whereas only  $30 \pm 9\%$  of Monologue participants and  $35 \pm 8\%$  of Controls uncovered at least three features (see Fig. 5A). Logistic regression yielded a significant effect of Communication (Dialogue condition vs. the two individual conditions) on the (binary) likelihood of uncovering three or more features,  $\chi^2(1) = 10.71, p = .001$ . The effect of Speaking (Monologue vs. Control) failed to reach significance,  $\chi^2(1) = 0.16, p = .69$ .

This pattern of selective attention mirrored the "likelihood of *mentioning* three or more features" when participants in the Dialogue and Monologue conditions referred to creatures during training. Over half ( $57 \pm 7\%$ ) of Dialogue participants mentioned at least three features during the last training block, but only  $15 \pm 8\%$  of Monologue participants met or exceeded that three-feature minimum (see Fig. 5B). Again, logistic regression on the (binary) likelihood of *mentioning* three or more features corroborated this difference,  $\chi^2(1) = 10.64, p = .001$ . In all, Dialogic participants distributed their attention more broadly than individuals.



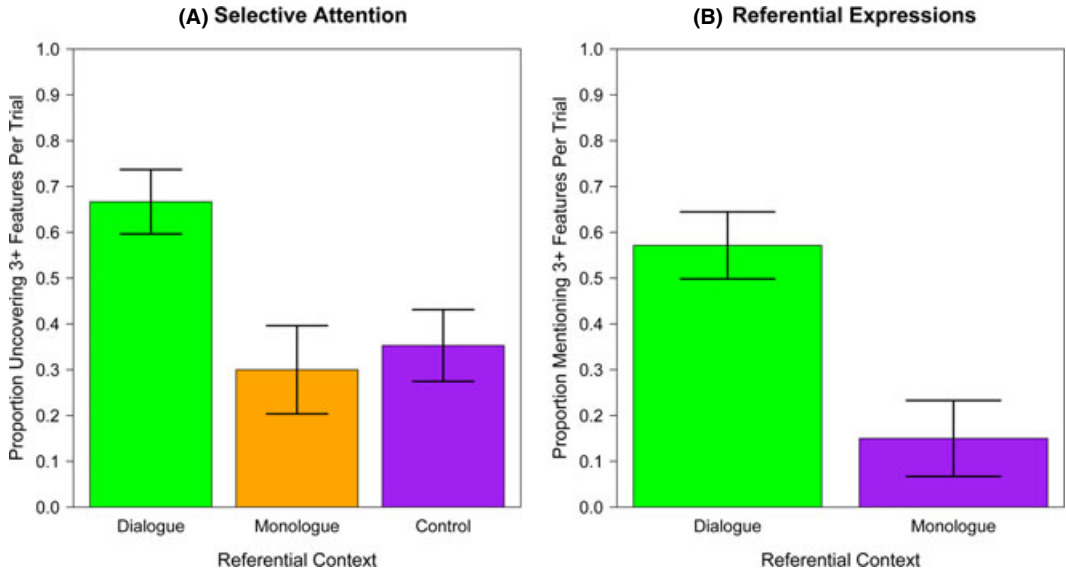


Fig. 5. (A) Comparison of the proportion of participants in the three referential conditions who *uncovered* three or more features per trial while predicting functions during the Attention Allocation posttest. (B) Comparison of the proportion of participants in the Dialogue and Monologue conditions who mentioned three or more features per trial while predicting functions during training. Error bars indicate 95% confidence interval.

### 3.2.2. Dialogue widens the distribution of attention across feature relations

To accurately predict the functional categories, participants could not simply attend severally to the various individual features; they needed to allocate attention to various relationships among features and between features and functions. Predicting the function associated with the family resemblance structure required attention to particularly complex relations. For example, in the category structure depicted by Table 1 (see Materials), the types of Tentacles, Fins, and Heart define a family resemblance structure that predicts whether the creature is nutritive. Physically uncovering or mentioning three or more features would imply attention to family resemblance, but the preceding analyses reveal little about the attentional tendencies of participants who uncovered or mentioned fewer than three features.

Here, we reinforce the implications of the preceding analyses by assessing the attentional tendencies related specifically to the family resemblance structure. We use the data on uncovered features to compute the “likelihood of uncovering family-resemblance features” (i.e., the proportion of *all possible* family-resemblance features that participants uncovered). Likewise, we compute the “likelihood of mentioning family-resemblance features” during the last block of Dialogic and Monologic training.

As was evident in the category learning results, Dialogue participants certainly predicted the family-resemblance function with greater accuracy than individual learners. Based on the posttest attention data, participants in the Dialogue condition also appeared more likely ( $79 \pm 4\%$ ) than either Monologue participants ( $64 \pm 5\%$ ) or Controls ( $57 \pm 5\%$ ) to uncover family-resemblance features when making their predictions (see Fig. 6A).

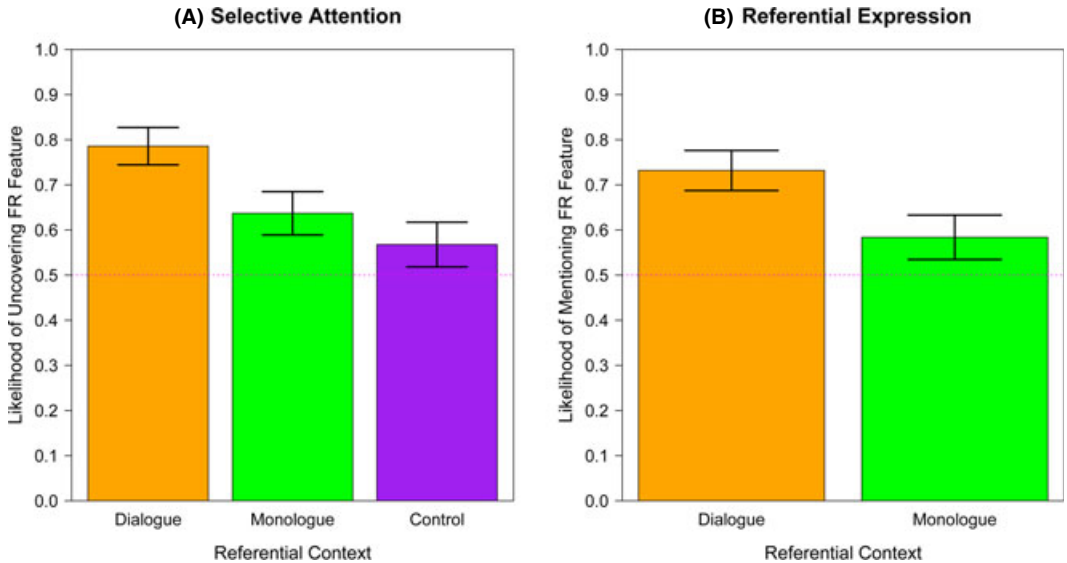


Fig. 6. (A) Comparison of the likelihood of participants in the three referential conditions *uncovering* features belonging to the family resemblance structure when predicting functions during the Attention Allocation post-test. (B) Comparison of the likelihood of participants in the Dialogue and Monologue conditions *mentioning* features belonging to the family resemblance structure when predicting functions during training. Error bars indicate 95% confidence interval.

An ANOVA on the (arcsine-transformed) “likelihood of uncovering family-resemblance features” underscored this advantage for the Dialogue condition. The effect of Communication (Dialogue vs. the other two conditions) was significant,  $F(1, 91) = 15.22, p < .001, \eta_p^2 = .14$ . The effect of Speaking (Monologue vs. Control) failed to reach significance, with  $\eta_p^2 < .01$ .

Again, reference in the Dialogue and Monologue conditions foreshadowed posttest patterns of selective attention (see Fig. 6B). Dialogic participants ( $73 \pm 4\%$ ) were significantly more likely than Monologue participants ( $58 \pm 5\%$ ) to mention family-resemblance features during the final training block,  $\chi^2 = 4.04, p = .04$ . Taken together, Dialogic participants appear to have attended to more feature/function relations than individuals.

That said, one must infer attention to feature/function relations from the likelihoods of uncovering or mentioning family resemblance features. The referential expressions recorded during Dialogic and Monologic training also provide direct evidence of attention to relations: verbal conjunctions. Dialogic partners described creatures using  $M = 1.035$  ( $SD = 0.792$ ) conjunctions—specifically, the words *and*, *but*, and *with(out)*—per training trial; participants in the Monologic condition never used conjunctions in their referring expressions. Overall, feature conjunctions appear more clearly noted and better learned through dialogue.

### 3.2.3. Linking patterns of reference to patterns of selective attention.

Throughout the Results on selective attention, we have presented referential results along with behavioral results (the features that participants physically uncovered before making a

prediction) and directed the reader's attention to how the former relate to the latter. While compelling, these side-by-side comparisons of aggregate patterns of behavioral and referential attention do not necessarily demonstrate a direct link between the two. That link becomes more evident when one considers the data on the sequential order in which participants uncovered features during the posttest and the order in which they mentioned features during training. We used that data to assess "attentional alignment": the extent to which temporal patterns of selective attention mirrored patterns of reference during both Dialogue and Monologue.

We derived our measure of attentional alignment by converting sequences of features uncovered during the posttest and sequences of features mentioned during the last training block into strings of feature codes—*T* (tentacles), *F* (fins), *H* (heart), *E* (eyes)—then calculated the edit distance (Damerau, 1964) between pairs of the 124 resulting strings (one reference string and one attention string for each of the 42 Dialogue participants and 20 Monologue participants). For example, if a participant mentioned *heart, fins, tentacles*, during a training trial, the resulting string was *HFTx* (we used an "x" for omitted features to remove length of string from the similarity measurement). Likewise, if the same participant uncovered *heart, tentacles, fins*, during a posttest trial, the resulting string was *HTFx*. The string-edit distance between what the participant mentioned and uncovered is 1 (Damerau's algorithm counts inversions as one edit rather than two). Actual strings consisted of 128 characters (four features by 32 trials). As elsewhere, we used two coders (one blind to the purpose of our research) to code the feature sequences. Agreement was high (Krippendorff's  $\alpha = .92$ ), and disagreements were resolved through discussion.

Patterns of selective attention did, in fact, align with patterns of reference. Reference strings were more similar (i.e., lower edit distance) to attention strings within participants ( $M = 8.07$ ,  $SD = 2.65$ ) than between participants ( $M = 9.51$ ,  $SD = 2.31$ ); this difference was significant,  $z = -4.89$ ,  $p < .001$ . In other words, participants in both verbal conditions uncovered features in more or less the same order in which they mentioned features, but in a different order from other participants.

This does not mean that Dialogic partners maintained their own idiosyncratic patterns of selective attention. In fact, attention strings were better aligned among Dialogic partners ( $M = 7.38$ ,  $SD = 2.55$ ) than among non-conversing pairs of Dialogic participants ( $M = 8.72$ ,  $SD = 1.87$ ),  $z = -4.26$ ,  $p < .001$ , or among pairs of Monologic participants ( $M = 9.22$ ,  $SD = 2.28$ ),  $z = -4.89$ ,  $p < .001$ .

Most important for the hypothesized link between referential processes, the attention strings of Dialogic participants were more similar to the reference strings of their partners ( $M = 8.33$ ,  $SD = 2.23$ ) than among randomly selected pairs of Dialogue participants who had not conversed with one another ( $M = 9.29$ ,  $SD = 2.57$ ),  $z = -2.32$ ,  $p = .02$ , or among randomly selected pairs of Monologic participants ( $M = 9.21$ ,  $SD = 2.55$ ),  $z = -2.05$ ,  $p = .04$ . At least when it comes to patterns of selective attention, interlocutors may indeed manipulate one another's conceptual processes.

### 3.2.4. Summary of selective attention (H2) results

As hypothesized, the results demonstrated that negotiating reference during Dialogue widened the distribution of attention across diagnostic features and yielded greater

awareness of how features relate to one another and to functions. Inventing a private lexicon during Monologue offered no benefits over working silently. In addition, we demonstrated a strong (possibly causal) link between the addressee's patterns of selective attention and the speaker's patterns of reference.

### 3.3. Information pooling: A competing hypothesis

Dialogue appears to generate an unambiguous category-learning advantage, with widely distributed attention to diagnostic features and structural relations between features and functions. We attribute this advantage to social-pragmatic processes. Others might point out simple quantitative differences between the three referential conditions. Dyadic interlocutors had access to two (likely disjoint) perspectives on how one might differentiate creatures into useful categories. Participants in the other two conditions had access to only one perspective, their own. Arguably, then, one might explain the category-learning advantages of Dialogic participants as a consequence of simply pooling the information gleaned from two perspectives rather than as a consequence of negotiating reference. To test this competing hypothesis, we devised two simulations of information pooling: one that simulates the aggregation of attentional information, a second that simulates the aggregation of structural information.

#### 3.3.1. Pooling attentional information

The Attention Allocation Posttest provided data on which features participants uncovered during each of the 32 prediction trials. We created 1,431 pseudo-dyads by pairing each of the 54 participants in both the Monologue and Control conditions with every other individual participant. For each pair, we then computed the union of features uncovered on each trial; for example, if one participant in a pseudo-dyad uncovered the creature's *fins* and *heart*, and the other participant uncovered the *tentacles* and *heart*, the union would equal *fins*, *heart*, *tentacles*. Averaging the number of features uncovered across trials and across pseudo-dyadic pairs, we find that information pooling yields attention to  $M = 2.78$  ( $SD = 0.39$ ) features per trial. By contrast, Dialogic participants uncovered  $M = 3.27$  ( $SD = 0.75$ ) features per trial; *actual* dialogue broadens attention significantly more than pseudo-dialogue,  $z = 7.77$ ,  $p < .001$  (see Fig. 7A).

#### 3.3.2. Pooling structural information

We, likewise, simulated the pooling of structural information. The Free Sort Posttest provided data on which creatures were grouped together by participants. As described in the analyses of the posttest sort data, we converted both the participant-created clusters of creatures and the four "true" categories into binary co-occurrence vectors. We then computed the union of sort-cluster vectors for each of 1,431 pseudo-dyads, and calculated the normalized mutual information (category fidelity) between these integrated co-occurrence vectors and the "true" co-occurrence vector. On average, information pooling yields category fidelity of  $M = 0.06$  ( $SD = 0.04$ ); Dialogue yielded much higher category fidelity,  $M = 0.38$  ( $SD = 0.43$ ). Again, *actual* dialogue improves category learning significantly more than pseudo-dialogue,  $z = 25.2$ ,  $p < .001$  (see Fig. 7B).

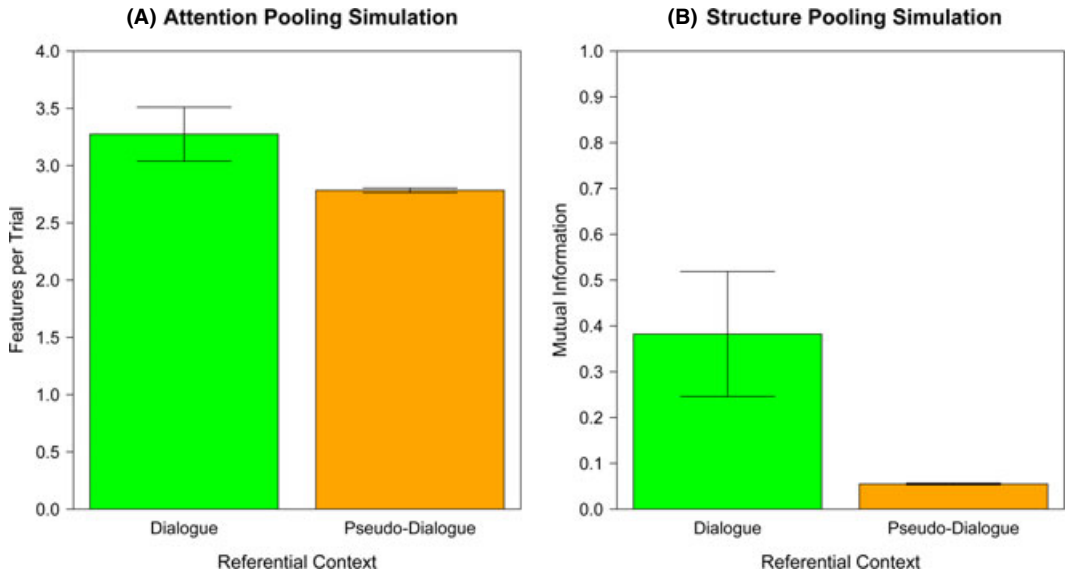


Fig. 7. Comparison of the Dialogic participants to pseudo-dyads (all pairs of individual participants) on (A) the number of features uncovered (we used the union of features uncovered for pseudo dyads) during the Attention Allocation posttest; and (B) the similarity (category fidelity) of the post-training sort clusters (we used the union of sort-clusters for pseudo-dyads) to the “true” category clusters. Error bars indicate 95% confidence interval.

### 3.3.3. Summary of information pooling

Thus, the simple (informational) pooling of perspectives does not appear to explain the category learning advantages of participants in the Dialogue condition. It is possible that some function other than the union of selective attention and/or perceived categories might better capture how Dialogic partners pool their perspectives, but we suspect otherwise. As argued and hypothesized, negotiating reference appears to compel interlocutors to uncover (both physically and metaphorically) more information than two people working in parallel. In the discussion, we introduce the notion of a *negotiated* coupling or integration of perspectives.

## 4. Discussion

This study was motivated by the notion that “public” and “private” uses of language yield different conceptual consequences. We presented participants with novel objects related by perceptual and functional features and asked which better facilitates category learning and use: negotiating shared reference during dialogue or inventing a private lexicon during monologue. As hypothesized, the results demonstrated that dialogue led to fast and accurate category learning, with widely distributed attention across diagnostic features, and greater awareness of how features relate to one another and to functions. Monologue offered no benefits over working silently.

Our results support theories that posit social-pragmatic constraints on (and, perhaps, origins of) referential and conceptual structures (Csibra & Gergely, 2009; Steels & Belpaeme, 2005; Tomasello, 2005). Conceptual structure emerges (or emerges more efficiently) as people share conceptual knowledge and, at the same time, to enable the sharing of such knowledge (Freyd, 1983). Social cognition—specifically, the skills interlocutors use to understand one another’s communicative intentions—plays an essential role in category learning (cf. Tomasello, 2000); neither garden-variety cognitive skills nor language use, in and of itself, appears to suffice for efficient learning of complex categories. That said, our results do not necessarily refute previously demonstrated effects of labeling objects on category-related learning, attention, and memory outside of explicit communication (e.g., Lupyan, 2008a, 2008b; Lupyan et al., 2007). In fact, our results also align with notions that language use can make conceptual processes more concrete (James, 1918/1950), manipulable (A. Clark, 2006), and durable (Dennett, 1994), providing the conceptual structure needed for judging subtle conjunctions (Gentner, 2003) and disjunctions (E. V. Clark, 1987) in a perceptually noisy environment. Nevertheless, our results suggest that these seemingly extracommunicative effects of language on conceptual processes do in fact derive from communication—both explicit and implicit.

As pointed out earlier, experimenters communicate implicitly with their subjects when they provide a lexicon for referring to experimental stimuli. In these conversational ultimatum games, the investigator’s lexicon serves as the focal point, the one salient solution to the task at hand (cf. Schelling, 1960). Subjects are compelled to look for what patterns among the objects explain the lexical distinctions (cf. Waxman & Markow, 1995). Explanations can reveal these underlying patterns (e.g., Williams & Lombrozo, 2009). One must wonder whether our monologic participants explained their own lexical choices while inventing their private lexicons. On the other hand, prior research has demonstrated that interlocutors constantly explain one another’s choices, at least until they have established reference and those explanations become expectations of future choices (Clark & Brennan, 1991). For example, to interpret an expression such as “this creature has sharp hands and rapidly pulsing heart, but short tentacles,” the beamer might wonder: why the “and” and “but;” why does the spotter relate these features? The observed advantages of dialogue over monologue may derive from such ongoing explanatory processes and expectations.

Dialogic partners in our study could not ask each other to explain lexical choices. Conversation was restricted to descriptions and requests for re-descriptions; they had to infer what was meant from what (little) we permitted them to say (cf. Clark & Lucy, 1975). Doing so required perspective taking (Krauss & Fussell, 1991); partners had to imagine what the other saw and believed as they produced and interpreted referring descriptions. Human beings appear especially motivated (perhaps hardwired; Tomasello, Carpenter, Call, Behne, & Moll, 2005) and skilled (again, perhaps hardwired; Herrmann, Call, Hernández-Lloreda, Hare, & Tomasello, 2007; Moll & Tomasello, 2007) to imagine one another’s mental states. Moreover, human beings tend to base judgments and decisions on one another’s beliefs much as they would their own (cf. Kovács, Téglás, & Endress, 2010). Even if such imaginings ran shallow (Keysar, Barr, Balin, & Brauner, 2000), efforts at perspective taking likely compelled partners to allocate attention to more features in more combinations than

either would have done on his or her own. Negotiating reference widened the distribution of attention to diagnostic features and yielded better learning of perceptual and functional feature conjunctions. By contrast, an egocentric perspective sufficed for inventing a private lexicon for a narrow, private conception of the task at hand.

This egocentrism of monologue may clarify certain aspects of the language-and-thought debate (see Gentner & Goldin-Meadow, 2003). For example, when performing a non-linguistic task such as judging whether a toaster is more similar to a man or woman (Boroditsky, Schmidt, & Phillips, 2003), one need not imagine the perspectives of anyone beyond one's self; one's native language—including the grammatical gender of toasters—may serve as the one focal point for making such private judgments. However, the second perspective conveyed through implied communication may explain contradictory findings in this debate, such as whether one's color lexicon constrains color categorization (again, see Gentner & Goldin-Meadow, 2003). Asking subjects to judge the similarity between colors may invite category distinctions they need to interpret the request, but usually ignore.

Perspective taking would also have permitted the coupling of reasoning processes despite our prohibitions against explicit reference to feature/function relationships. The results of the information-pooling (attention and structure) simulations rule out a simple aggregation of conceptual information; Dialogic learning exceeds the union of two perspectives. Instead, negotiating reference may yield a *negotiated* integration of perspectives, specifically the integration of relational hypotheses. That is, partners could interpret one another's descriptions as hypotheses about how to allocate attention not only to features but, as suggested by the use of conjunctions in Dialogic expressions, how to allocate attention to feature combinations when predicting functions. Consequently, partners could couple their diverse perspectives and test a richer pool of hypotheses (cf. Wiley & Jolly, 2003), especially relational hypotheses. Moreover, one might expect that interdependence between partners would motivate this more subtle form of cognitive coupling (Deutsch, 1949; Johnson & Johnson, 1989). The spotter's score depended on the beamer's predictions, which, in turn, depended on the spotter's descriptions. This interdependence would compel the beamer to go beyond the mere information given in the spotter's descriptions; in fact, the mere belief that one is collaborating with another person (as opposed to a computer program) appears enough to compel such conceptual leaps (cf. Okita, Bailenson, & Schwartz, 2007). Even if partners were wrong about one another's perspectives, the coupling of one's own perspective with that of an imagined other may yield a richer perspective. In the present study, Dialogic participants certainly outperformed individuals.

That said, some might continue to argue that the observed dialogic advantages reflect nothing more than social facilitation (cf. Zajonc, 1965) or increased motivation due to the aforementioned interdependence of Dialogic participants. The data alone cannot refute such an argument; we can only discuss how we tried to control the possible motivational differences between referential contexts. We tried to motivate individual learners by offering a prize (a digital music player) to the best performing participant(s). The prize was certainly salient to participants, given that everyone asked about it while being debriefed. Further, we tried to take advantage of audience effects (i.e., the presence of the investigator) and co-action effects (participating at the same time with another participant) on individual motivation. The

adequacy of these motivators remains open. Beyond these controls, we also argue that greater motivation is far from guaranteed in collective activity. For example, collaborators are prone to social loafing (Latané, Williams, & Harkins, 1979), expecting their partners to take up any slack in their own efforts. Even in the absence of loafing, collaborators might choke under the pressure to perform for the benefit of their partners (cf. Ariely, Gneezy, Loewenstein, & Mazar, 2009). Another problem is poor rapport; following the experiment, some Dialogic participants complained about uncooperative interlocutors. We could not link these debriefing responses to performance, but one might suspect that these complaints came from poorly performing dyads. In all, motivation is not a binary factor; one cannot simply argue that motivation was “on” in the Dialogue condition and “off” in the individual conditions.

Along the same lines, category learning occurs in more or less collaborative (and interdependent) settings, where learners rely more or less on social-pragmatic processes. In the present study, we considered differences in category learning at the polar extremes: heavy reliance on social-pragmatic reasoning during dialogue; little if any during monologue. Much remains between these extremes for us (and others) to explore (see Noveck & Reboul, 2008; Noveck & Sperber, 2004, for current directions in experimental pragmatics). Research on remote collaboration (e.g., Kraut, Fussell, & Siegel, 2003) demonstrates one direction in which to take future research: start with collaborators in a dialogic setting and decrease social-pragmatic cues step by step until they perform no better than isolated individuals. Alternatively, we prefer an approach whereby we start with an isolated individual and, step by step, increase social pragmatic cues. Research on the emergence of human communication systems (e.g., Galantucci, 2005; also Scott-Phillips, Kirby, & Ritchie, 2009) and on collaboration with artificial agents (e.g., Okita et al., 2007) both suggest ways in which to manipulate the social pragmatics of category learning.

## 5. Conclusion

As things stand, our findings further the understanding of how communication and language use, in general, influence conceptual processes. We have demonstrated that negotiating conventions of reference during dialogue enhances category discovery. Inventing a private lexicon during monologue does not. Dialogue may yield a subtle coupling of reasoning processes by compelling interlocutors to imagine one another’s perspectives as they explain and mirror one another’s lexical choices. This understanding can inform investigations into innumerable human endeavors that depend on communication.

## Notes

1. In this way, the single player would encode messages when speaking his or her monologue and decode those messages when listening to those messages.
2. Dialogic *spotters* and *beamers* resemble *speakers* and *addressees* or *directors* and *matchers* in other related research. We avoid those terms because, unlike previous



research, the roles of *spotter* and *beamer* also apply to individual participants, whereas the other terms do not.

3. We use an arcsine transformation of these (function- and category-prediction accuracy) and several subsequent proportional values when performing analyses of variance (ANOVAS; Anscombe, 1948).
4. The responses of dialogic partners were dependent on one another during training; thus, the dyad serves as the unit of analysis in the first and second ANOVA. Elsewhere, we analyze individual data.

## Acknowledgments

This research derives in part from the doctoral dissertation submitted by John Voiklis to Columbia University. The dissertation committee members—Deanna Kuhn, Peter T. Coleman, Robert O. McClintock, and Oded Koenigsberg—provided helpful guidance. We thank Gregory Murphy and Doris Zahner for providing thoughtful feedback on earlier drafts of this report.

## References

- Anscombe, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35(3/4), 246–254.
- Ariely, D., Gneezy, U., Loewenstein, G., & Mazar, N. (2009). Large stakes and big mistakes. *Review of Economic Studies*, 75, 1–19.
- Austin, J. L. (1975). *How to do things with words*. Cambridge, MA: Harvard University Press.
- Barr, D. J., & Kronmüller, E. (2006). Conversation as a site of category learning and category use. In A. B. Markman & B. H. Ross (Eds.), *Psychology of learning and motivation: Categories in use* (Vol. 47, pp. 181–211). San Diego, CA: Academic Press.
- Boroditsky, L., Schmidt, L. A., & Phillips, W. (2003). Sex, syntax, and semantics. In D. Gentner & S. Goldin-Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 61–79). Cambridge, MA: MIT Press.
- Brennan, S. E. (1995). Centering attention in discourse. *Language and Cognitive Processes*, 10(2), 137–167.
- Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482–1493.
- Brennan, S. E., Galati, A., & Kuhlen, A. K. (2010). Two minds, one dialog: Coordinating speaking and understanding. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 53, pp. 301–344). San Diego: Elsevier.
- Brown, R. W. (1958). How shall a thing be called. *Psychological Review*, 65(1), 14–21.
- Brown, J., Aczel, B., Jiménez, L., Barry, S., & Plaisted, K. (2010). Intact implicit learning in autism spectrum conditions. *Quarterly Journal of Experimental Psychology*, 63(9), 1789–1812.
- Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language*, 61(2), 171–190.
- Chiu, C.-Y., Krauss, R. M., & Lau, I. Y. M. (1998). Some cognitive consequences of communication. In S. R. Fussell & R. J. Kreuz (Eds.), *Social and cognitive approaches to interpersonal communication* (pp. 259–278). Hillsdale, NJ: Erlbaum.

- Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. In B. Macwhinney (Ed.), *Mechanisms of language acquisition* (pp. 1–33). Hillsdale, NJ: Lawrence Erlbaum.
- Clark, H. H. (1996). *Using language*. Cambridge, UK: Cambridge University Press.
- Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences*, 10(8), 370–374.
- Clark, E. V., & Amaral, P. M. (2010). Children build on pragmatic information in language acquisition. *Language and Linguistics Compass*, 4(7), 445–457.
- Clark, A., & Karmiloff-Smith, A. (1993). The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind & Language*, 8(4), 487–519.
- Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). Washington, DC: American Psychological Association.
- Clark, H. H., & Lucy, P. (1975). Understanding what is meant from what is said: A study in conversationally conveyed requests. *Journal of Verbal Learning and Verbal Behavior*, 14(1), 56–72.
- Clark, H. H., & Murphy, G. L. (1982). Audience design in meaning and reference. In J. F. LeNy & W. Kintsch (Eds.), *Advances in psychology, Vol. 9. Language and comprehension* (pp. 287–299). Amsterdam: North-Holland.
- Cruse, D. A. (1977). The pragmatics of lexical specificity. *Journal of Linguistics*, 13(2), 153–164.
- Csibra, G., & Gergely, G. (2009). Natural pedagogy. *Trends in Cognitive Sciences*, 13(4), 148–153.
- Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM*, 7(3), 171–176.
- Dennett, D. C. (1994). The role of language in intelligence. In J. Khalfa (Ed.), *What is intelligence? The Darwin College Lectures*. (pp. 161–178) Cambridge, UK: Cambridge University Press.
- Dessalegn, B., & Landau, B. (2008). More than meets the eye: The role of language in binding and maintaining feature conjunctions. *Psychological Science*, 19(2), 189–195.
- Deutsch, M. (1949). A theory of co-operation and competition. *Human Relations*, 2(2), 129–152.
- Duff, M. C., Hengst, J. A., Tranel, D., & Cohen, N. J. (2005). Development of shared information in communication despite hippocampal amnesia. *Nature Neuroscience*, 9(1), 140–146.
- Echterhoff, G., Lang, S., Krämer, N., & Higgins, E. T. (2009). Audience-tuning effects on memory. *Social Psychology*, 40(3), 150–163.
- Freyd, J. J. (1983). Shareability: The social psychology of epistemology. *Cognitive Science*, 7(3), 191–210.
- Fussell, S. R., & Krauss, R. M. (1989a). The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology*, 25(3), 203–219.
- Fussell, S. R., & Krauss, R. M. (1989b). Understanding friends and strangers: The effects of audience design on message comprehension. *European Journal of Social Psychology*, 19(6), 509–525.
- Galantucci, B. (2005). An experimental study of the emergence of human communication systems. *Cognitive Science*, 29, 737–767.
- Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: A study in conceptual and semantic coordination. *Cognition*, 27(2), 181–218.
- Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, 1(2), 292–304.
- Gastgeb, H. Z., Strauss, M. S., & Minshew, N. J. (2006). Do individuals with autism process categories differently? The effect of typicality and development. *Child Development*, 77(6), 1717–1729.
- Gelman, S. A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition*, 23(3), 183–209.
- Gentner, D. (2003). Why we're so smart. In D. Gentner & S. G. Meadow (Eds.), *Language in mind: Advances in the study of language and thought* (pp. 195–235). Cambridge, MA: MIT Press.
- Gentner, D., & Goldin-Meadow, S. (Eds.). (2003). *Language in mind: Advances in the study of language and thought*. Cambridge, MA: MIT Press.

- Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics 3: Speech acts* (pp. 41–58). New York: Academic Press.
- Hengst, J. A. (2003). Collaborative referencing between individuals with aphasia and routine communication partners. *Journal of Speech, Language, and Hearing Research*, 46(4), 831–848.
- Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science*, 317(5843), 1360–1366.
- Holtgraves, T. M., & Kashima, Y. (2008). Language, meaning, and social cognition. *Personality and Social Psychology Review*, 12(1), 73–94.
- James, W. (1950). *The principles of psychology*. New York: Dover. (Original work published 1918)
- Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and competition: Theory and research*. Edina, MN: Interaction Book Company.
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological Science*, 11(1), 32–38.
- Knowlton, B. J., & Squire, L. R. (1993). The learning of categories: Parallel brain systems for item memory and category knowledge. *Science*, 262(5140), 1747–1749.
- Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, 330(6012), 1830–1834.
- Krauss, R. M. (1987). The role of the listener: Addressee influences on message formulation. *Journal of Language and Social Psychology*, 6(2), 81–98.
- Krauss, R. M., & Fussell, S. R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition*, 9(1), 2–24.
- Krauss, R. M., Vivekananthan, P. S., & Weinheimer, S. (1968). "Inner speech" and "external speech": Characteristics and communication effectiveness of socially and nonsocially encoded messages. *Journal of Personality and Social Psychology*, 9(4), 295–300.
- Krauss, R. M., & Weinheimer, S. (1966). Concurrent feedback, confirmation, and the encoding of referents in verbal communication. *Journal of Personality and Social Psychology*, 4(3), 343–346.
- Kraut, R. E., Fussell, S. R., & Siegel, J. (2003). Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction*, 18(1), 13–49.
- Krippendorff, K. (1986). *Information theory: Structural models for qualitative data*. Thousand Oaks, CA: Sage.
- Latané, B., Williams, K., & Harkins, S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology*, 37(6), 822–832.
- Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.
- Lupyan, G. (2008a). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition*, 108(2), 566–577.
- Lupyan, G. (2008b). From chair to "chair": A representational shift account of object labeling effects on memory. *Journal of Experimental Psychology: General*, 132(2), 348–369.
- Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, 18(12), 1077–1083.
- Malt, B. C., & Sloman, S. A. (2004). Conversation and convention: Enduring influences on name choice for common objects. *Memory and Cognition*, 32(8), 1346–1354.
- Malt, B. C., Sloman, S. A., & Gennari, S. P. (2003). Universality and language specificity in object naming. *Journal of Memory and Language*, 49(1), 20–42.
- Markman, A. B., & Makin, V. S. (1998). Referential communication and category acquisition. *Journal of Experimental Psychology: General*, 127(4), 331–354.
- Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, 129(4), 592–613.
- Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences*, 34(2), 57–74.
- Minda, J. P., & Ross, B. H. (2004). Learning categories by making predictions: An investigation of indirect category learning. *Memory & Cognition*, 32(8), 1355–1368.

- Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 639–648.
- Mueller Gathercole, V. C., Cramer, L. J., Somerville, S. C., & Jansen op de Haar, M. (1995). Ontological categories and function: Acquisition of new names. *Cognitive Development*, 10(2), 225–251.
- Noveck, I. A., & Reboul, A. (2008). Experimental pragmatics: A Gricean turn in the study of language. *Trends in Cognitive Sciences*, 12(11), 425–431.
- Noveck, I. A., & Sperber, D. (2004). *Experimental pragmatics. Palgrave studies in pragmatics, languages and cognition*. New York: Palgrave Macmillan.
- Okita, S. Y., Bailenson, J., & Schwartz, D. L. (2007). The mere belief of social interaction improves learning. In D. S. McNamara & J. G. Trafton (Eds.), *The proceedings of the 29th meeting of the cognitive science society* (pp. 1355–1360). Austin, TX: Cognitive Science Society.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02), 169–190.
- Schelling, T. C. (1960). *The strategy of conflict*. Cambridge, MA: Harvard University Press.
- Schober, M. F., & Clark, H. H. (1989). Understanding by addressees and overhearers. *Cognitive Psychology*, 21(2), 211–232.
- Schober, M. F., & Conrad, F. G. (1997). Does conversational interviewing reduce survey measurement error? *The Public Opinion Quarterly*, 61(4), 576–602.
- Scott-Phillips, T. C., Kirby, S., & Ritchie, G. R. S. (2009). Signalling signalhood and the emergence of communication. *Cognition*, 113(2), 226–233.
- Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science*, 13(1), 13–19.
- Soulières, I., Mottron, L., Saumier, D., & Larochelle, S. (2007). Atypical categorical perception in autism: Autonomy of discrimination? *Journal of Autism and Developmental Disorders*, 37(3), 481–490.
- Sperber, D., & Mercier, H. (in press). Reasoning as a social competence. In H. Landemore & J. Elster (Eds.), *Collective wisdom*. Cambridge, UK: Cambridge University Press.
- Stalnaker, R. C. (2002). Common ground. *Linguistics and Philosophy*, 25(5–6), 701–721.
- Steels, L., & Belpaeme, T. (2005). Coordinating perceptually grounded categories through language: A case study for colour. *Behavioral and Brain Sciences*, 28(4), 469–489.
- Steels, L., & Kaplan, F. (2002). AIBO's first words: The social learning of language and meaning. *Evolution of Communication*, 4(1), 3–32.
- Tomasello, M. (2000). The social-pragmatic theory of word learning. *Pragmatics*, 10(4), 401–414.
- Tomasello, M. (2005). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28(5), 675–691.
- Vygotsky, L. (1986). *Thought and language*. Cambridge, MA: MIT Press. (Original work published 1962)
- Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, 29(3), 257–302.
- Wiley, J., & Jolly, C. (2003). When two heads are better than one expert. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the twenty-fifth annual conference of the cognitive science society* (pp. 1241–1245). Hillsdale, NJ: Erlbaum.
- Wilkes-Gibbs, D., & Clark, H. H. (1992). Coordinating beliefs in conversation. *Journal of Memory and Language*, 31(2), 183–194.
- Williams, J. J., & Lombrozo, T. (2009). Explaining promotes discovery: Evidence from category learning. In N. A. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31th annual conference of the cognitive science society* (pp. 1186–1191). Austin, TX: Cognitive Science Society.
- Wittgenstein, L. (2001). *Philosophical investigations: 50th anniversary commemorative edition* (G. E. M. Anscombe and E. Anscombe, Trans.). Malden, MA: Blackwell Publishing. (Original work published 1958)
- Zajonc, R. B. (1965). Social facilitation. *Science*, 149(3681), 269–274.