April 29 2010

Conventional Wisdom: Negotiating Conventions of Reference Enhances Category Learning

John Voiklis
Stevens Institute of Technology
James E. Corter
Teachers College, Columbia University

UNCORRECTED DRAFT, PLEASE DO NOT CITE OR QUOTE
Published version available from *Cognitive Science*, DOI: 10.1111/j.1551-6709.2011.01230.x

*Keywords: reference, pragmatics, joint activity, coordination, category learning*

Correspondence concerning this article should be addressed to John Voiklis, 211 West 107th Street, #1W, New York, NY 10025. Email: voiklis@gmail.com
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 shareability 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
Conventional Wisdom: Negotiating Conventions of Reference Enhances Category Learning

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). 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 (Brennan & Clark, 1996; also Garrod & Anderson, 1987). 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, through subsequent repetition, 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 consider the conceptual effects of referential communication and arrive at the Shareability Hypothesis (at least, our interpretation), which posits communicative constraints on (and, perhaps, origins of) conceptual structure (Freyd, 1983). We then contrast explicit processes of dialogic reasoning to the implicit processes of individual reasoning and hypothesize that shareability constraints on dialogic processes 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 have no special advantages over nonverbal individual 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.

**On the relationship between referential and conceptual processes**

Referential communication involves 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 (cf., Barr & Kronmüller, 2006; Brown, 1958; Cruse, 1977). We do not argue here that lexical choices determine how collaborators categorize objects to which they have already established reference (cf., Malt, Sloman & Gennari, 2003). Rather, lexical choices signal social-pragmatic design (e.g. deixis, presupposition, implicature, etc.): speakers design referential expressions for interpretation and addressees interpret those expressions assuming design (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 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 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) -- e.g., 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., Wilkes-Gibbs & Clark, 1992; Krauss, 1987; Krauss & Weinheimer, 1966). In summary, the apprentice manipulates the chef’s conceptual processes by designing a referring expression that directs the chef’s attention towards 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.
On the relationship between referential and conceptual structure

What conceptual utility accrues from such a conceptual pacts? Conceptual pacts appear to stabilize the conceptual effects of negotiating reference. Adhering to conceptual pacts helps interlocutors more easily direct joint attention (e.g., Richardson, Dale, & Kirkham, 2007), confirm joint construal (cf., Clark & Krych, 2004), and execute joint intentions (cf., Clark & Lucy, 1975). More importantly, following conversation, conceptual pacts influence how each collaborator sorts the objects 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. The Shareability Hypothesis (Freyd, 1983) posits that conceptual structure -- specifically categorical structure -- emerges as people share conceptual knowledge and to enable the sharing of such knowledge; i.e. people represent (at least, socially-derived) knowledge in a shareable form. Therefore, we next ask whether negotiating reference with a collaborator also 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 conceptual processes and strategies from those outside of conversation (cf., 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, S. 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 (Soulières, Mottron, Saumier & Larochelle, 2007; Gastgeb, Strauss & Minshew, 2006). As observable in connectionist models (cf., A. Clark & Karmiloff-Smith, 1993), implicit reasoning yields structures that often defy verbalization (required for shareability), 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 (cf., Sperber & Mercier, in press; Mercier & Sperber, in press). We do not doubt that typical individuals can discern the same structures in their private activities as collaborators discern in their public activities (cf., Malt, Sloman & Gennari, 2003), but individual processes may take more time and yield less differentiated structure than dialogue.

Some conceptual consequences of labeling categories

That said, similar conceptual effects might derive from simply verbalizing conceptual processes without explicit communication (cf., Vygotsky, 1986 [1962]). 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, focuses attention on how perceptual features relate to one another (Gentner, 1993) or to functional features (Mueller Gathercole, Cramer, Somerville, & Jansen op de Haar, 1995) and helps 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 contrasting use of the two labels were intended to communicate. Similarly, Lupyan (2008b) instructed subjects to label objects using basic-level terms -- e.g., “chair” as opposed to “recliner.” Basic-level terms convey
socially-salient information about the referent (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]).

**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 (Krauss, Vivekananthan & Weinheimer, 1968; Fussell & Krauss, 1989a, 1989b). We contend, though, that private referential expressions might convey conceptual content but that content lacks (shareable) structure: 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. Private reference or monologue functions as a one-sided conversation and, consequently, exhibits minimal interpretable design (Krauss et al., 1968; Fussell & Krauss, 1989a, 1989b). Consequently, we doubt that monologue yields conceptual utility comparable to dialogue.

**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 \([H3]\) yield more shareable referential expressions and \([H2]\) 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, 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. We implemented the task as a computer game, in which players predict whether various alien creatures provide food (±nutritive) and/or present a threat (±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). 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.

**Method**

**Participants**

Ninety-six male and fifty 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 four years post-secondary schooling during which they devoted approximately two hours per week to computer games.

**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 contrasts of these conditions. Communication contrasted

---

1 In this way, the single player would encode messages when speaking their monologue, and decode those messages when listening to those messages.
dialogic participants against individuals (both Monologic participants and Controls). *Speaking* contrasted Monologic participants against Controls.

**Materials**

Participants repeatedly classified sixteen 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 90-110 percent of its standard measurement.

Insert Table 1 here

Participants classified creatures on two functionally significant features (criterion variables): whether or not a creature produced jelly usable as food and fuel (±nutritive) and whether or not it might damage life support systems (±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 or not the creature is nutritive. A value of “1” on at least two of these three features means the creature is nutritive, otherwise not. Eye-type can function as a simple rule to predict whether or not the creature is destructive. 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 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.

Procedure & Tasks

*Pretest.* Before training, all participants individually performed a *Free Sort* of the sixteen 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 sixteen creatures randomly arranged in a single container. While in a container, creatures were rendered at 15% of their normal size in order 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 one to sixteen) as they deemed necessary. The pretest free sort provided data on the *number of features mentioned* by participants when explaining creature groupings.

______________________________________________________

*Insert Figure 1 here*

______________________________________________________
**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: ten blocks of thirty-two trials (i.e. two random 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 beamer² as they collaborated through networked workstations divided by a 5’x 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 the 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).

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.
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, one of sixteen creatures appeared at the center of the spotter’s video display, where it remained until the beamer acted or twenty seconds (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. In order 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, as well as the number and order of features to which participants referred in the Monologue and Dialogue conditions.
**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 by graphical blinds (Fig. 1C shows the Attention Allocation interface). Participants literally *uncovered* (mouse-clicked 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.

**Dependent Measures and Analyses**

Our somewhat novel research paradigm (a hybrid of psycholinguistics and categorization methodologies) and use of novel versions of traditional tasks (e.g., explanations of sorting behavior), yields complex data that requires uncommon methods of analysis. Thus, it seems helpful to first describe how we derived the various dependent measures from the data. We describe more common measures and analyses in the Results section.

**Deriving measures of category learning.** We derived dependent measures of category learning from the Function-prediction Training task and the Posttest Free Sort task. The training task provided data on correct and incorrect predictions of individual functions (±nutritive and
predictions were coded “1” when correct and “0” when incorrect. We derive our measure of **function-prediction accuracy** by computing the proportion of correct predictions within each of the ten blocks of thirty training trials. One can also 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 the ten blocks. We use an arcsine transformation of these proportional values (both function and category-prediction) when performing analyses of variance (Anscombe, 1948).

The Posttest Free Sort data provides more 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. Our measure of **category avowal** derives from the average proportion of category citations across sort clusters. Again, we use an arcsine transformation of the proportional category avowal values when performing analyses of variance.

**Deriving measures of selective attention.** We derived separate dependent measures of selective attentional patterns from the Attention Allocation Posttest and from the referring expressions recorded during Dialogic and Monologic training. These sources provided data on the number of overall and family-resemblance features that participants either (physically) uncovered and/or mentioned when predicting functions.

In order to predict *both* functions with 100% accuracy, participants needed to uncover a minimum of three features: one for the function predictable by a simple unidimensional rule, and at least two out of the three family-resemblance features that probabilistically predict the other function. So, we computed the **likelihood of uncovering three or more features** (i.e., “1” for participants who uncovered 3+ features and “0”) and the **likelihood of uncovering family-resemblance features** (i.e., the proportion of *all possible* family-resemblance features that participants uncovered). Likewise, using the referring expressions recorded during the last block of training, we computed the **likelihood of mentioning three or more features** and the **likelihood of mentioning family-resemblance features**. We use logistic regression to model the binary measures of uncovering/mentioning three or more features. As elsewhere, we use an arcsine transformation of the proportional likelihoods of uncovering and mentioning family-resemblance features when performing analyses of variance.

Additionally, the Attention Allocation Posttest and reference recordings provide data on the order of features that participants uncovered and/or mentioned when predicting functions. A
speaker can manipulate the conceptual processes of an addressee by designing referring expressions that direct the addressee's attention towards features that both differentiate the target from other possible referents and allow the addressee to infer the referent's significance to the activity. We infer this phenomenon from similarities between patterns of features uncovered and features mentioned.

Specifically, we converted the features that each participant uncovered during the posttest and the features the each participant mentioned during the last block of thirty-two training trials into strings of feature codes -- T (tentacles), F (fins), H (heart), E (eyes) -- then derived our measure of **attentional alignment** by calculating the edit distance (Damerau, 1964) between pairs of the 124 resulting strings (one reference string and one attention string for each of the forty-two Dialogue participants and twenty 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 thirty-two trials). As before, 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.

**Deriving measures of shareability.** Negotiating reference compels collaborators to find communicable structure in their shared activity. Shareability, then, connotes the extent to which the design of referring expressions reflects that communicable structure. Shareability can also
connote the *effort expended* in designing shareable expressions: **length of expression**, **complexity of expression**, and **type of expression**. We derived dependent measures from the referential expressions recorded during Dialogic and Monologic training.

The length of referring expressions (average word count) seems an intuitive measure of design effort. Also, word count has been widely used in psycholinguistic research (cf., Carroll, 1980). We measured **length of expression** by computing the number of words participants used per trial.

One might also infer design effort from the grammatical complexity of referring expressions (diversity of grammatical cases; see Krauss & Weinheimer, 1967; Krauss & Glucksberg, 1977, on the complexity of social and nonsocial speech). To measure **complexity of expression**, two coders (again, one blind to the purpose of our research) classified each word in the referential data as a *noun*, an *adjective*, or *other* (*Krippendorff’s alpha* = .98). We chose these grammatical cases, because one could minimally describe any of the binary features with a noun and/or adjective. For example, one could describe the two types of fins as “fins” vs. “flippers” or “spiky” vs. “webbed” or “spiky fingers” vs. “webbed fingers.” The proportion of words per trial belonging to grammatical cases other than nouns and adjectives served as an indicator of the effort expended on referential complexity.

Collaborators actively monitor one another’s conceptual (cf., Clark & Krych, 2004) and procedural (cf., Bangerter & Clark, 2003) activities. An individual “collaborating” with his or her future self must, more or less, do the same. Messages that monitor these activity-related processes serve as our final indicator of the effort expended on designing shareable expressions. The verbal data was transcribed by professional transcription service that specializes in psychological interviews and protocols. Using punctuation and other discourse markers, we
segmented the verbal data into phrases. Two coders (one blind to the purpose of our research) then classified each phrase into two **types of expressions**: those that referred to mental and procedural processes and those that did not. Agreement, again, was high (Krippendorff’s alpha = .88), and disagreements were resolved through discussion. We measured this aspect of design using the proportion of words per trial devoted to talk about process.

**Results**

We have organized the results around our three hypotheses: [H1] dialogue enhances category learning; [H2] dialogue broadens attention to features and yields greater attention to have features relate to functions; and [H3] dialogue increases the shareability of expressions. 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. In support of Hypothesis 3, we report analyses of several measures of the effort participants expended when designing referential expressions. We conclude by presenting some additional analyses of an alternative hypothesis: specifically we ask whether information pooling might explain the wider distribution of Dialogic attention and better category learning.

**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 allows us to compare category-learning **efficiency** (how quickly participants appear to learn the experimenter-defined category structure). The sort data allows us to compare category-learning **efficacy** (how well participants ultimately learn that structure).
**Dialogue yields more efficient category learning**

**Category-prediction accuracy.** As described in Dependent Measures and Analyses we inferred learning of the four implicit categories (defined by the conjunction of the two functions, ±nutritive and ±destructive) by measuring when participants correctly predicted the conjunction both functions for stimulus creatures. We derived category-prediction accuracy by averaging correct predictions of function conjunctions. We computed this measure for each block of thirty-two trials.

---

Insert Figure 2 here

---

The advantage for Dialogic dyads\(^3\) is evident in Fig. 2, which compares category-prediction accuracy by referential context across the 320 training trials (ten blocks of thirty-two trials). Dialogic dyads predicted functions with 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 between-groups contrasts defined in the **Design**: *Communication* (comparison of Dialogue versus Monologue and Control) and *Speaking* (comparison of Monologue versus Control), with Block as a within-subjects factor. The

\(^3\) 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.
interaction of Communication and Block was significant -- $F(9, 648) = 14.6594, p < .001 \eta^2_p = .17$ -- as were the main effects of Communication -- $F(1, 72) = 33.86, p < .001, \eta^2_p = .32$ -- and Block -- $F(9, 648) = 38.87, p < .001, \eta^2_p = .35$. Neither the main effect of Speaking nor the interaction of Speaking and Block reached significance -- both, $\eta^2_p < .01$. Across training trials, accuracy increased more quickly in the Dialogue condition than in the individual learning conditions.

---

**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 features (family resemblance) features that predict the other function. 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 derive our measure of **function-prediction accuracy** by computing the proportion of correct predictions (separately for the simple-rule function and family-resemblance function) within each of the ten blocks of thirty 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 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 \textbf{function-prediction accuracy} values). The effects of Communication were significant in a three-way interaction with Structure-Type (family resemblance versus 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 ($F(1,72) = 2.99, p = 0.09, \eta_p^2 = .04$). Dialogic participants predicted simple-rule and family-resemblance with comparable overall accuracy. In fact, 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 versus 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 ($\eta_r^2 < .01$), in a three-way interaction with Structure-Type and Block ($\eta_r^2 = .01$), or in interactions with either Structure-Type or Block (both $\eta_r^2 < .01$). 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.

*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 direct evidence of whether participants learned categories. As described in Dependent Measures and Analyses, the Posttest Free Sort provided data on which creatures grouped into the same sort clusters by participants. We measured **category fidelity** as the normalized mutual information (i.e. similarity) between the sort clusters created by participants and the four “true” (experimenter-designed) implicit categories. The Posttest Free Sort also provided data on whether participants cited observable features, individual functions, or categories (conjunctions of functions or predictive behaviors, such as *capture & stun*, etc.) to justify creature groupings. 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.
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 -- $M$(dialogue) = .38 ($SD = .43$) vs. $M$(monologue) = .06 ($SD = .08$) and $M$(control) = .09 ($SD = .14$). 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% ±16% (dialogue) vs. 26% ±16% (monologue) and 28% ±16% (control). Category fidelity correlated significantly with category avowal, $r = .52, p < .001$.

A 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 -- $F(1, 93) = 11.89, p < .001, \eta^2_p = .21$ -- and significant univariate effects on category fidelity -- $F(1, 93) = 24.05, p < .001, \eta^2_p = .21$ -- and on category avowal -- $F(1, 93) = 5.79, p = .02, \eta^2_p = .06$. Neither the multivariate nor univariate effects of Speaking on category fidelity or on category avowal reached significance, with all $\eta^2_p < .01$. In all, dialogic participants, more than individuals, explicitly sorted in accordance with the “true” category clusters.

**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.

**H2. Dialogue widens selective attention across features and feature relations**

Early in this paper, we argued that interlocutors manipulate one another’s conceptual processes by designing referring expression that direct the addressee’s attention towards 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 point to one mechanism (among other potential mechanisms) through which negotiating reference can facilitate category learning (cf., Garrod & Pickering, 2009).

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.

**Dialogue widens selective attention to features**

In order to predict both functions with 100% accuracy, participants needed to uncover a minimum of three features: one for the function predictable by a simple unidimensional rule, and at least two out of the three features that probabilistically predict the other function. Further,
uncovering at least three features might imply that participants have learned to distinguish the four categories implicitly defined by the two functions. As described in Dependent Measures and Analyses, we computed the **likelihood of uncovering three or more features** in each referential condition.

During the individual posttest, 67% ±7% of Dialogue participants met or exceeded that three-feature minimum, while only 30% ±9% of Monologue participants and 35% ±8% of Controls uncovered at least three features (see Fig. 5A). Logistic regression yielded a significant effect of Communication (Dialogue condition versus the two individual conditions) on the (binary) likelihood of uncovering three or more features, $B = .47, p = .001$. The effect of Speaking (Monologue versus Control) failed to reach significance, $B = .12, p = .69$.

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% ±7%) of Dialogue participants mentioned at least three features during the last training block, but only 15% ±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, $B = 2.02, p < .01$. In all, Dialogic participants distributed their selective attention more broadly than individuals.

**Dialogue widens selective attention to 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 or not the creature is nutritive. A value of “1” on at least two of these three features means the creature is nutritive, otherwise not.

As 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% ±4%) than either Monologue participants (64% ±5%) or Controls (57% ±5%) to uncover family-resemblance features when making their predictions (see Fig. 6A).

An ANOVA on likelihood of uncovering family-resemblance features (i.e., the arcsine-transformed proportion of all possible family-resemblance features that participants uncovered) underscored this advantage for the Dialogue condition. The effect of Communication (Dialogue versus the other two conditions) was significant -- $F(1, 91) = 15.22, p < .001, \eta_p^2 = .14$. The effect of Speaking (Monologue versus 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% ±4%) were significantly
more likely than Monologue participants (58% ±5%) to mention family-resemblance features during the final training block – *chi-squared* = 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* = .792) conjunctions -- specifically, the words *and*, *but*, and *with(out)* -- per training trial; participants in the Monologue condition never used conjunctions in their referring expressions. Overall, feature conjunctions appear more clearly noted and better learned through dialogue.

*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 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 patterns of selective attention mirrored patterns of reference during both Dialogue and Monologue.

As described in Dependent Measures and Analyses, 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 between pairs of the 124 resulting strings (one reference string and one attention string for each of the forty-two Dialogue participants and twenty 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 (inversions count as one edit rather than two). Actual strings consisted of 128 characters (four features by thirty-two trials).

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 the 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.

**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.

**H3. Inferring shareability from differences in Dialogic and Monologic expressions**

A comparison of the referring expressions used by participants to describe stimuli under conditions of Dialogue and Monologue should yield insights into whether and how communicating with a partner differs from communicating with oneself. Specifically, we looked for differences in shareability (i.e., communicable structure or design). Shareability was evident in the analyses of **attentional alignment** and conjunction usage reported in the results related to the selective attention hypothesis. Shareability, here, connotes the **effort expended** in designing referential expressions: **length of expression**, **complexity of expression**, and **type of expression** (as defined in **Dependent Measures and Analyses** and, briefly, below). Naturally, participants in the Control condition generated no verbal behavior to analyze.
**Inferring shareability from length of expression**

The length of referring expressions has long been acknowledged as an indicator of social and nonsocial design (Carroll, 1980). For private, self-directed speech, people tend towards minimal design, using elliptic referring expressions (Krauss & Glucksberg, 1977; cf. also Matsuka & Corter, 200x). For public, other-directed speech, people often start with longer, sometimes prolix, expressions, then shorten and simplify those expressions with repeated reference (Krauss & Weinheimer, 1964). Both early wordiness and later brevity signal social-pragmatic design. In all, we expected the word count of referring expressions to be initially higher and decrease more sharply in the Dialogue condition than in the Monologue condition.

Indeed, across the first and last blocks of training (trials 1-32 and 289-320, respectively), Dialogic dyads used a greater number of words per trial than did Monologic participants -- $M_{(dialogue)} = 11.08$ (SD = 4.33) vs. $M_{(monologue)} = 6.50$ (SD = 2.97) (see Fig. 7A). This main effect of Dialogue vs. Monologue on length of referring expression was confirmed by a repeated-measure ANOVA, $F(1,60) = 41.75, p < .001, \eta^2_p = .41$. Nevertheless, while participants in both referential contexts reduced the number of words per trial from the first training block -- $M_{(dialogue)} = 13.33$ (SD = 4.71) vs. $M_{(monologue)} = 7.43$ (SD = 3.39) -- to the last training block -- $M_{(dialogue)} = 8.82$ (SD = 2.30) vs. $M_{(monologue)} = 5.57$ (SD = 2.18) -- this reduction was greater in the Dialogue condition, as confirmed by the test of interaction with training block, $F(1,60) = 4.58, p = .04, \eta^2_p = .07$.

**Inferring shareability from complexity of expression**

Similarly, referring expressions also differed in complexity between referential contexts. Dialogic dyads used complex expressions, such as “this creature has sharp hands and rapidly pulsing heart, but short tentacles.” Expressions like these include words belonging to a variety of
grammatical cases. Monologic participants used simple expressions constructed mainly of adjectives and nouns, such as “flashy middle, even legs.” As described in Dependent Measures and Analyses, we used the proportion of words per trial belonging to grammatical cases other than nouns and adjectives to measure **complexity of expression**: i.e., the amount of effort expended on referential design (cf., Krauss & Weinheimer, 1967; Krauss & Glucksberg, 1977). Half (50% ±4%) of the words per description used by Dialogic dyads were not nouns or adjectives; Monologic participants used fewer words per description from other cases (17% ±7%). This asymmetry in grammatical complexity was significant—chi-squared = 214.14, p < .001 (see Fig. 7B).

**Inferring shareability from type of expression**

Activity-related messages, both about the processes of the training task (cf., Bangerter & Clark, 2003) and about the processes of learning (cf., Clark & Krych, 2004), can serve as an additional indicator of the greater design effort in Dialogue over Monologue. This might seem an unfair comparison: remote partners must communicate about one another’s contributions the joint activity, but individuals seem unlikely to monitor their own processes out loud. Nevertheless, Monologic messages were meant for one’s future self; as training progressed, it would seem reasonable to expect participants in the Monologue condition to leave themselves messages about their future activities. Using the method described in Dependent Measures and Analyses, we segmented the verbal data into phrases then sorted those phrases into two types of expressions: those that referred to mental and procedural processes and those that did not.

As expected, Dialogic dyads devoted slightly more than half (54% ±4%) of the words per trial to discussing the process of their collaborative learning (see Fig. 7C). Dyadic communication included utterances related to turn taking (e.g., “now, I’m spotting”), utterances
relating the state of the game (e.g., “the energy meter is running low”) or the state of their learning (e.g., “I think we’ve figured it out”), and several exclamations and expletives. As further expected, Monologic participants also commented on process (especially expletives and assessments of their learning, such as “this is hard”) in the recordings they made for themselves, though their comments entailed a smaller proportion of words per trial (24% ±7%) than Dialogic comments. This asymmetry in process-monitoring messages was significant -- $\text{chi-squared} = 150.85, p < .001$.

Summary of shareability (H3) results

In all, negotiating reference elicits (perhaps requires) more conceptual effort than establishing reference with oneself. Much of that effort appears to go into designing shareable exchanges, as evident in both the length of dialogic expressions and the eventual foreshortening of those expressions. Dialogic partners also expended more effort than Monologic participants on designing expressions with greater grammatical diversity and devoted more lexical effort discussing and monitoring their learning processes. That extra conceptual effort likely yielded the observed conceptual advantages for Dialogue over Monologue.

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, including the effort expended on designing shareable expressions. Others might point out simple quantitative
differences between the three referential conditions. Dyadic interlocutors had access to (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. In order 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.

**Pooling attentional information**

The Attention Allocation Posttest provided data on which features participants uncovered during each of the thirty-two prediction trials. We created 1,431 pseudo-dyads by pairing each of the fifty-four 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 = .39) features per trial. Dialogic participants uncovered $M = 3.27$ (SD = .75) features per trial; actual dialogue broadens attention significantly more than pseudo-dialogue, $z = 7.77, p < .001$ (see Fig. 8A).
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 Dependent Measures and Analyses, 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 $M = .06$ (SD = .04); Dialogue yielded much higher category fidelity, $M = .38$ (SD = .43). Again, actual dialogue improves category learning significantly more than pseudo-dialogue, $z = 25.2$, $p < .001$ (see Fig. 8B).

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 perceived categories might better capture how Dialogic partners pool their perspectives, but we suspect otherwise. In the discussion, we introduce the notion of a negotiated coupling or integration of perspectives.

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 (Tomasello, 2005; Steels & Belpaeme, 2005; also Csibra & Gergely, 2009). Conceptual structure emerges (or emerges more efficiently) as people share conceptual knowledge and to enable the sharing of such knowledge (Freyd, 1983). Social-cognition -- 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, appear 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 judgements 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 selective attention to diagnostic features and yielded better
learning of perceptual and functional feature conjunctions. In contrast, an egocentric perspective sufficed for inventing a private lexicon.

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 or not 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 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). 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 step-by-step decrease social-pragmatic cues 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, 2008) and on collaboration with artificial agents (e.g., Okita, Bailenson & Schwartz, 2007) both suggest ways in which to manipulate the social pragmatics of category learning.

Conclusion

As things stand, our findings further the understanding of how communication and language use, in general, influence conceptual processes. We 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.
References


*Biometrika*, 35(3/4).


Acknowledgments

We thank Greg Murphy and Doris Zahner for providing thoughtful feedback.
Table 1. The category structure used in the experiment. Illustrates one of four possible assignments of diagnostic surface features (tentacles, fins, heart, eyes) to category structure.

<table>
<thead>
<tr>
<th>EXEMPLAR</th>
<th>TENT.</th>
<th>FINS</th>
<th>HEART</th>
<th>EYES</th>
<th>COLOR</th>
<th>SIZE</th>
<th>NUTR.</th>
<th>DEST.</th>
<th>CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td>ND.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>random</td>
<td>1</td>
<td>1</td>
<td>ND</td>
</tr>
<tr>
<td>ND.2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>random</td>
<td>1</td>
<td>1</td>
<td>ND</td>
</tr>
<tr>
<td>ND.3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>random</td>
<td>1</td>
<td>1</td>
<td>ND</td>
</tr>
<tr>
<td>ND.4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>random</td>
<td>1</td>
<td>1</td>
<td>ND</td>
</tr>
<tr>
<td>Nx.1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>random</td>
<td>1</td>
<td>0</td>
<td>Nx</td>
</tr>
<tr>
<td>Nx.2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>random</td>
<td>1</td>
<td>0</td>
<td>Nx</td>
</tr>
<tr>
<td>Nx.3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>random</td>
<td>1</td>
<td>0</td>
<td>Nx</td>
</tr>
<tr>
<td>Nx.4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>random</td>
<td>1</td>
<td>0</td>
<td>Nx</td>
</tr>
<tr>
<td>xD.1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>random</td>
<td>0</td>
<td>1</td>
<td>xD</td>
</tr>
<tr>
<td>xD.2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>random</td>
<td>0</td>
<td>1</td>
<td>xD</td>
</tr>
<tr>
<td>xD.3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>random</td>
<td>0</td>
<td>1</td>
<td>xD</td>
</tr>
<tr>
<td>xD.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>random</td>
<td>0</td>
<td>1</td>
<td>xD</td>
</tr>
<tr>
<td>xx.1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>random</td>
<td>0</td>
<td>0</td>
<td>xx</td>
</tr>
<tr>
<td>xx.2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>random</td>
<td>0</td>
<td>0</td>
<td>xx</td>
</tr>
<tr>
<td>xx.3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>random</td>
<td>0</td>
<td>0</td>
<td>xx</td>
</tr>
<tr>
<td>xx.4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>random</td>
<td>0</td>
<td>0</td>
<td>xx</td>
</tr>
</tbody>
</table>
Figure Captions.

Figure 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 participant uncover (remove graphical blinds) a creature feature-by-feature before predicting its function.

Figure 2. Compares 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.

Figure 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 mid-point of the training task (trial 160).

Figure 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.

Figure 5. A) Compares 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) Compares 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.

Figure 6. A) Compares the likelihood of participants in the three referential conditions uncovering features belonging to the family resemblance structure when predicting functions during the Attention Allocation posttest. B) Compares 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.

Figure 7. A) Compares the length of expressions (words per trial) used by participants in the Dialogue and Monologue conditions during the first and last blocks (each 32 trials) of training. B) Compares the complexity of expressions (words other than adjectives and nouns) used by participants in the Dialogue and Monologue conditions when referring to creatures during training. C) Compares the proportion of activity-monitoring words (type of expressions) used by participants in the Dialogue and Monologue conditions during training. Error bars indicate 95% confidence interval.

Figure 8. Compares 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.
A. Function Prediction

B. Pre/Post Free Sort

C. Attention Allocation
Figure 2.

![Graph showing the category prediction accuracy across different function prediction trials.

- **Dialogue**: Black circles.
- **Monologue**: Triangles.
- **Control**: Grey squares.

The x-axis represents the function prediction trials, while the y-axis shows the category prediction accuracy. The graph demonstrates an increasing trend in accuracy for all categories as the trials progress.
Figure 3.

A. Simple Rule

B. Family Resemblance
Figure 4.

A. Mutual Information

B. Explanations
Figure 5.

### A. Selective Attention

- **Proportion Uncovering 3+ Features Per Trial**

<table>
<thead>
<tr>
<th>Referential Context</th>
<th>Dialogue</th>
<th>Monologue</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### B. Referential Expressions

- **Proportion Mentioning 3+ Features Per Trial**

<table>
<thead>
<tr>
<th>Referential Context</th>
<th>Dialogue</th>
<th>Monologue</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6.
Figure 7.
Figure 8.