New Ideas in Psychology 30 (2012) 179–189

Contents lists available at SciVerse ScienceDirect

New Ideas in Psychology

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

Conceptual agreement theory

Sergio E. Chaigneau [a,c,*], Enrique Canessa [b,c], Joaquín Gaete [a]

[a] Escuela de Psicología, Universidad Adolfo Ibáñez, Avenida Diagonal Las Torres 2640, Peñalolén, Santiago, Chile
[b] Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Balmaceda 1625, Recreo, Viña del Mar, Chile
[c] Centro de Investigación de la Cognición – CINCO, Universidad Adolfo Ibáñez, Avenida Diagonal Las Torres 2640, Peñalolén, Santiago, Chile

Keywords:
Reference
Shared reference
Meaning
Agreement
Joint action

A B S T R A C T

For some time now, psychological inquiry on reference has assumed that reference is achieved through causal links between words and entities (i.e., direct reference). In this view, meaning is not relevant for reference or co-reference. We argue that this view may be germane to concrete objects, but not to diffuse objects (that lack clear spatio-temporal limits, thus preventing the use of direct reference in interactions). Here, we propose that meaning is the relevant dimension when referring to diffuse entities, and introduce Conceptual Agreement Theory (CAT). CAT is a mathematized theory of meaning that specifies the conditions under which two individuals (or one individual at two points in time) will infer they share a diffuse referent. We present the theory, and use *stereotype stability* and *public opinion* as case studies to illustrate the theory's use and scope.

1. Introduction

When using a concept to refer to something (e.g., please pass the *jelly*), we want the addressee to select the correct referent, though meaning and lexical entry may not completely coincide (e.g., understand the utterance as a request for the *marmalade*). Consistent with Putnam's (1973) view that meaning and reference are separate, much research now shows that people use direct reference (e.g., pointing, gazing), rather than meaning, to develop agreement in situations like these. However, what happens when people talk about a *diffuse referent*? By this, we mean reference to process-like entities that lack well defined spatio-temporal limits and that generate intersubjective and even intrasubjective differences in the way they are construed (e.g., some abstract entities, social processes and institutions, emotional states, self identity). How can people agree, if they cannot point to a diffuse referent? The view we develop here is that agreeing about diffuse referents is all about meaning. We assume that when people agree about a diffuse something, the question people try to answer is how does someone else understand a diffuse something, which would explain her acting in such and such a way or saying such and such a thing.

Obviously, given our emphasis in meaning, we need to offer a theory of it. The reader will find it in a later section. In the next section, we develop in greater detail the idea of diffuse referents and their relation to reference in general.

2. Diffuse referents

For almost 40 years now, many philosophers have agreed that problems of meaning and problems of reference are separate. Following Quine's (1951) point that meaning does not have a one to one correspondence to sense-data, and extending on views advanced by Kripke (1980), Putnam (1973) developed a causal theory of reference. In this theory, meaning cannot solve the problem of reference, which is achieved via causal relations between an object and the act of asserting it. Names and natural kinds, in this view, are rigid designators linked to their referent by a history of reference. A classical example of this

* Corresponding author. Escuela de Psicología, Universidad Adolfo Ibáñez, Avenida Diagonal Las Torres 2640, Peñalolén, Santiago, Chile. Tel.: +56 2 331 1199.

E-mail addresses: sergio.chaigneau@uai.cl (S.E. Chaigneau), ecanessa@uai.cl (E. Canessa), joaquin.gaete@uai.cl (J. Gaete).

0732-118X/$ – see front matter
doi:10.1016/j.newideapsych.2011.09.001

is christening someone as, say, “Laura”. After Laura is christened, the act of naming her is transmitted to other people in a causal chain that allows her to be named Laura even by people who were not present in the original christening. In this theory, names are associated to objects' spatio-temporal identity and not to their meaning. Reference in the Kripke/Putnam view, then, has two important characteristics. First, it relies on direct reference, not on meaning. Second, for natural kinds, there are experts (e.g., scientists) in society in which we rely to recognize a kind (i.e., there is a linguistic division of labor).

A view like this has been enormously influential in cognitive psychology. Though philosophical concerns generally focus on epistemic issues (e.g., Quine, 1969), and psychological concerns generally focus on sharing reference in language, conversation and in joint action (see Galantucci & Sebanz, 2009), the general idea that reference requires establishing causal links between objects and actions has been profusely exploited in psychology. The types of causal links studied are things such as gaze and joint attention (e.g., Moses, Baldwin, Rosicky, & Tidball, 2001; Richardson, Dale, & Tomlinson, 2009; Tomasello, 1995), eye movements (e.g., Spivey, Tanenhaus, Eberhard, & Sedivy, 2002), pointing gestures (e.g., Carpenter, Nagell, & Tomasello, 1998), and conceptual pacts (e.g., Brennan & Clark, 1996; Garrod & Anderson, 1987).

Note, however, that all the preceeding work is mainly relevant when making reference to concrete objects. Evidently, much has been learned by focusing on cases such as how do people communicate when their task is to manipulate concrete objects (e.g., Brown-Schmidt & Tanenhaus, 2008; Clark & Krych, 2004). Nevertheless, this focus leaves out instances of reference where the referent is *diffuse* (e.g., some abstract entities, social processes and institutions, emotional states, self identity). For two reasons, these diffuse referents are not amenable to the Putnamian analysis. First, direct reference to them is not an option in conversation (regardless of whether you think that it is merely impractical or that it is in principle impossible). Second, there are no experts in society to whom we can resort to recognize a true instance of a diffuse entity.[1]

In contrast to cognitive psychology, where conversation and joint action are often cast as a problem of precisely determining which entity is being referred to, research on pragmatics frequently focuses on certain kinds of indefinitions in conversation. That is the whole point, for example, of the study of *presuppositions*, which are linguistic structures that specify their meaning appealing to an utterance's context and that would remain ambiguous without it (see Levinson, 1983). What is even more, vagueness is at times considered a tool in conversation (e.g., Channell, 1994; Sperberg & Wilson, 1991), acting to provide focus and guiding the conversational partner's allotment of cognitive resources (e.g., using conversational devices like “sort of” moves the focus of conversation out of the precise definition of the instance being referred; see Jucker, Smith, & Lüdge, 2003).

Though the term *diffuse referent* has not been used previously, we see it in line with this tradition, and believe it already is the topic of a whole line of research on pragmatics derived from studies of face to face interaction in relatively unconstrained situations (e.g., Billig, 1996; Garfinkel, 1967; Sacks, 1972; Shoter, 1993). More recently, and regarding more institutionalized situations (e.g., Atkinson & Heritage, 1984; McHoul & Rapley, 2001; Van der Howen, 2009), many authors examine conversations in psychotherapy about what we here call diffuse referents. Examples are: psychological problems (Antaki, 2007; Davis, 1986), emotions (Edwards, 1997; Leudar et al., 2008), self identity and change (Arístegui et al., 2009; Gergen, 1991; Strong, Busch, & Couture, 2008). It is apparent that when people are talking about these things, they act as if they were referring to a concrete object (i.e., talking about something). Therefore, and unless we want to restrict reference only to spatio-temporally defined objects, we need a theory of reference to diffuse objects.[2]

This whole article, in consequence, is about actions and utterances that cannot be said to be felicitous by appealing to direct reference; and hence, it is about cases where shared reference must involve agreement in meaning.

3. Our (minimalist) theory of meaning

3.1. Property frequency distributions

Concepts can be defined by property frequency distributions. Perhaps the simplest task designed to study concepts is the property generation task[3] (e.g., Hampton, 1979; Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976, Exp. 1). In this task, people are required to list all properties that they can think of for a given concept. These lists are generally believed to provide conceptual content. When these properties are coded for property types, and accumulated for a large sample of individuals, one obtains a frequency distribution like the one shown in Fig. 1, Panel *C*. Some properties are frequently mentioned, other properties are mentioned less frequently, and other properties are seldom mentioned. This is part of the basic information to be found in category norms (e.g., McRae, Cree, Seidenberg, & McNorgan, 2005).

Generally, when people learn a category, there are one or more contrast categories available (Tversky, 1977). It is widely believed that categorizing an entity as an instance of a focal category, implies that in some metric (e.g., similarity) the entity is closer to the focal category than to the contrast category or categories (Nosofsky, 1986, 1987;

[1] This may be thought of as the standard interpretation of Plato's Dialogs.

[2] This is not an ontological discussion about whether things such as democracy, identity, happiness, or the like exist outside the minds of people, but rather a discussion about how people could achieve agreement when referring to these things, given that they can't point to them.

[3] Also know as the “feature listing task”, but we adhere to calling it “property listing”, for reasons discussed in Santos, Chaigneau, Simmons, and Barsalou (2011) and also in Simmons, Hamann, Harenski, Hu, and Barsalou (2008).

		i						
C	participants	a	b	c	d	e	f	g
	1	1	1	1	0	0	--	--
	2	1	1	1	0	0	--	--
	3	1	0	1	1	1	--	--
	4	1	1	0	1	0	--	--
	5	1	1	0	0	1	--	--
	frequencies	5	3	3	2	2	--	--
Cn	participants							
	1	--	--	--	1	1	0	1
	2	--	--	--	1	1	1	1
	3	--	--	--	0	1	1	1
	4	--	--	--	0	0	1	1
	5	--	--	--	0	0	1	1
	frequencies	--	--	--	2	3	4	5

Fig. 1. A simple conceptual structure with 2 concepts *C* and *Cn*, defined by two frequency distributions of properties (*i*), illustrating frequency and inter-property correlations. Dashes signal conceptual properties that are not part of participant's conceptual content for a specific concept.

Rosch & Mervis, 1975). The metric is senseless unless focal and contrast categories share part of their conceptual content (i.e., properties). Thus, if we did the property listing task simultaneously on a focal and corresponding contrast category, we would get two overlapping frequency distributions where a given property type's frequency depends on the (focal or contrast) concept under consideration (as illustrated in Fig. 1, Panels *C* and *Cn*), where properties *a* through *e* belong to concept *C* and properties *d* through *g* belong to concept *Cn*, and properties *d* and *e* are in the intersection of those two concepts.

Another feature of these distributions is that for appropriately contrasting categories, they exhibit an inter-property correlational pattern. Properties that are listed more frequently for focal (contrast) than for contrast (focal) categories, are also positively correlated. Frequency and inter-property correlations are part of the basic data that many categorization theories try to account for (e.g., Medin, Altom, Edelson, & Freko, 1982; Murphy & Wisniewski, 1989). Our theory of meaning does not need many more assumptions besides holding that there exists something like a concept in individual minds, which explains the pattern of frequencies and inter-property correlations that we have been talking about. Whatever your theory about what concepts are, it should at a very minimum explain these things.

3.2. Conventionality

A second assumption of our theory of meaning is that the lists of properties discussed above are conventional. By conventionality we mean that people in a social group share a common set of properties, and also share to a certain extent the conceptual structure that makes these properties cohere. Note that convention in this sense is not normative but descriptive. It says that there is a limited number of property types that can be listed in a given social group for a given concept. This kind of stability explains why category norms are useful for designing experiments, even across relatively long periods of time (for examples, see Van Overschelde, Rawson, & Dunlosky, 2004). Conventionality in this very broad descriptive sense is an overarching assumption of much research on cognition. Concepts, scripts, stereotypes, are all conventional in the sense described above (for illustrations, see Murphy, 2002). We remain neutral here about how these conventions are developed in human groups: if due to requirements of coordination or communication (e.g., Lewis, 1969, 1975), due to the reproduction of action (e.g., Millikan, 2005), or to the observation of regularities in the natural and social world (e.g., Berlin, 1992).

3.3. Intersubjective differences

One last and critical assumption that we make is that for a given situation, concepts are probably not homogeneous across a population, meaning that different people conceptualize the same situation differently. More importantly, differences may exist even when people conceptualize a situation similarly. Though similar construals will have a background of common properties, there will be intersubject differences and perhaps even intrasubject differences in time. A simple source of non-homogeneity is learning. When concepts are learned in natural environments (in contrast to experimental environments), the most likely situation is that people will be exposed to different learning sets, and so they will develop different versions of the same concept. This might be even more so if the concept being learned is a diffuse referent.

This is an important but potentially problematic assumption. As many scholars have noted since Frege, it is not trivial to assert that two people possess two versions of the same concept, and not simply two different concepts altogether (for philosophical discussion, see Frege, 1893; Glock, 2009; Russell, 1997). In the literature of concepts, inter and intrasubjective differences in conceptual content have also been considered problematic. Barsalou (1987, 1993) argues that if conceptual content is unreliable, then this suggests that concepts are not fixed entities in the mind, and that what people exhibit are categorization abilities instead. In contrast to viewing this as a problem that would lead us to ask questions about the nature of conceptualization, we take intersubjective and even intrasubjective differences as a given. And, we adopt a pragmatics point of view in assuming that two individual concepts are the same if someone cannot discriminate between those who deploy the concept based on how they use it on a given occasion (cf., Horwich, 1998, in particular, the chapter on Meaning as Use).

4. Conceptual Agreement Theory (CAT)

4.1. Conceptual formulation

As discussed earlier, there are several differences between referring to concrete objects versus diffuse entities. The following difference is the most important for us here. Of course, when utterances refer to concrete objects, reality will act as a kind of judge of whether agreement was achieved or not. However, when the referent is diffuse, the recourse to direct reference is either impractical or straightforwardly not possible (though people may operate as if it were).

CAT assumes that when people agree about a diffuse something, what is relevant is that they act as if they were of the same mind, and not so much whether they are in fact correct or not about some state of affairs. CAT assumes that the question people try to answer is how does someone else understand a diffuse something, which would explain her acting in such and such a way or saying such and such a thing. What concerns us here is the agreement about the interpretation of an action (cf., Garrod & Pickering, 2004).

The central tenet of CAT is the following: In a specific situation, a given focal conceptualization is probabilistically more consistent with a set of properties, and probabilistically less consistent with another set (though this last set could still reflect reasonable options within the situation at hand). When solving the problem of co-reference to a diffuse referent, people cannot have access to other people's mind, and so agreement about co-referred conceptualizations needs to be inferred. CAT assumes that conceptual agreement (disagreement) is inferred when someone (the Observer, *O*) in a situation construes a conceptualization that defines it, making some properties coherent (not coherent) with it, and then someone else (the Actor, *A*) in that situation conveys one of the coherent properties (one of the incoherent properties), enabling *O* to infer that their conceptualizations are shared (not shared).

As with any inference, inferred agreement can be wrong. True agreement would occur when *O* infers a shared conceptualization, and *A* in fact construed a version of the same conceptualization. Illusory agreement would occur when *O* infers a shared conceptualization, and *A* in fact construed an alternative conceptualization. Because focal and contrast conceptualizations are only probabilistically related to the properties that people can use to make these inferences, true and illusory agreement are also probabilistic affairs. The probability of true agreement is the probability that *O* witnesses *A* doing or uttering *i*, and *i* is consistent with *O*'s conceptualization, given that *A* in fact construed a version of the same conceptualization. Similarly, the probability of illusory agreement is the probability that *O* witnesses *A* doing or uttering *i*, and *i* is consistent with *O*'s construal given that *A* in fact construed an alternative conceptualization. Importantly, from *O*'s point of view, illusory and true agreement would look exactly the same.

Thinking about diffuse referents in this way, highlights another important point. Whereas some joint action literature has emphasized the achievement of common goals as a measure of co-reference (e.g., Clark & Krych, 2004), this is not necessarily the most important consequence of sharing reference. It is very likely that agreement is important for the sustained effort toward a common goal, even more than it is for the actual achievement of that goal (Marsh, Richardson, & Schmidt, 2009). Furthermore, agreement may occur even in the absence of a common goal. People often interpret other people's actions, and this may have consequences beyond the achievement of common goals. For example, someone's action may lead me to feel empathy or feel disgust toward that person, depending on the nature of my interpretation of the reasons for her actions. Though inferences about agreement may be illusory (e.g., *O* agrees with *A*'s closing the window because it is windy outside, but *A* actually closed it because it was noisy outside), it is likely that the momentary feeling of agreement has consequences for the dyad's interaction. All these phenomena could be accommodated in CAT.

4.2. Mathematical formulation

Lets assume that in a given situation, there is a conventional focal *C* conceptualization that applies, and also an alternative contrast *Cn* conceptualization (or conceptualizations) that might apply too (e.g., *left-wing* and *right-wing*; *feminine* and *masculine*; *party*, *social gathering*, or *reunion*) (as Fig. 2 illustrates).

Let,

O = an observer.
A = an actor.
i = A property type available in the situation, consistent with one or more conceptualizations of the situation.
C = A set of concepts that reside in individual minds across a population, that are all conventionally applicable in the situation, and that share property types to a degree that allows calling them versions of the same concept. *C* is defined by a probability distribution of its respective property types (i).
k_1 = The number of property types in *C*.
s_1 = The average number of property types coherent with concept *C* in an individual's mind ($s_1 \leq k_1$).
Cn = A set of alternative conceptualizations (as defined above for *C*) that could be applied to the given situation. *Cn* is defined by a probability distribution of its respective property types (i), which is different from the probability distribution for *C*.
k_2 = The number of property types in *Cn*.
s_2 = The average number of property types coherent with concept *Cn* in an individual's mind ($s_2 \leq k_2$).
u = The number of property types that are consistent with *C* and *Cn* (i.e., the cardinality of the $C \cap Cn$ set).

Recall, from our discussion in the CAT Conceptual Formulation section, that agreement occurs when *O* is in her version of *C*, and *A* utters or does *i*, with *i* being consistent with *O*'s version of *C*. This event is true agreement ($a1$) if *A* is truly in *C*, and it is illusory agreement ($a2$) when *A* is truly in *Cn*. For these two events ($a1$ and $a2$), we can compute probabilities as explained below.

4.3. Probability of true agreement, p($a1$)

Assume a finite universe of k_1 (1, 2, 3, …, k_1) property types, distributed with given probabilities, all of which are non-zero probability properties ($p(i) > 0$) (as in Fig. 2). Lets take n_c (1, 2, 3, …, n_c) samples (i.e., one for each version of *C*

C	samples	a	b	c	d	e	f	g
	s_{11}	1	1	1	0	0	--	--
	s_{12}	1	1	1	0	0	--	--
	s_{13}	1	0	1	1	1	--	--
	…							
	s_{1n}	1	1	0	0	1	--	--

s_1 = average number of properties coherent with concept C in an individual's mind.
k_1 = number of properties in C (in this case, 5).

Cn	samples				$C \cap Cn$			
	s_{21}	--	--	--	1	1	0	1
	s_{22}	--	--	--	1	1	1	1
	s_{23}	--	--	--	0	1	1	1
	…							
	s_{2n}	--	--	--	0	0	1	1

s_2 = average number of properties coherent with concept Cn in an individual's mind.
k_2 = number of properties in Cn (in this case, 4).

Fig. 2. A simple conceptual structure with 2 concepts C and Cn, defined by two frequency distributions, illustrating frequency and inter-property correlations, where participants are viewed as samples of conceptual content.

in the social group), of fixed size $s_1 > 1$, and with $s_1 \le k$, where s_{1n} is the nth sample of size s_1. Assuming that all samples are possible (though with different probabilities), then the total number of possible samples is

$$n_c = \frac{k_1!}{(k_1 - s_1)!s_1!} \quad (1)$$

Clearly, these n_c samples will have their own probability distribution, so

$$\sum_{i=1}^{n_c} p(s_{1i}) = 1 \quad (2)$$

Imagine now that sample s_{1o} is chosen from C as reference sample (i.e., O's sample) and then that another s_{1a} sample is chosen also from C for comparison (i.e., A's sample). Imagine also that in this second s_{1a} sample, one property i is chosen at random (i.e., whatever A decides to say or do in a specific situation). The true agreement probability (p(a1)) is the probability that this i property is also contained in the s_{1o} reference sample (i.e., that O sees it as evidence in favor of her own conceptualization). Computing p(a1) involves calculating the expectation that 2 such samples show a coincidence.

To formulate this probability, we need to define a coincidence operator © where $s_{1o} © s_{1a}$ = number of coincidences between the elements of sample s_{1o} and those of sample s_{1a}. Bear in mind that each sample is one of the n_c combinations of properties possible in C.

$$p(a1) = E\left[\frac{s_{11}©s_{11}}{s_1} + \frac{s_{11}©s_{12}}{s_1} + \ldots + \frac{s_{1nc}©s_{1nc}}{s_1}\right] \quad (3)$$

Two things are worthy of noting in Eq. (3). Note that it entails counting all coincidences found when comparing pairwise all possible property combinations of size s_1. Note also that dividing by s_1 assumes that properties coherent with A's version of concept Cn, have all an equal probability of being offered as evidence for O to consider. We acknowledge that this last assumption is perhaps an oversimplification, justified only as a means to reduce the complexity of our analysis. However, we believe that relaxing this assumption would not substantially alter our subsequent claims.

Now, precisely because s_1 is a common factor in all quotients in (3), we can factor s_1 out of the expectation in (3). If we also take the expectation relative to the probability of the successive s_{1a} and s_{1o} samples, then we have

$$p(a1) = \frac{1}{s_1}\sum_{o=1}^{nc}\sum_{a=1}^{nc} s_{1o}©s_{1a}p_o p_a \quad (4)$$

where p_o = probability of sample s_{1o}, and p_a = probability of sample s_{1a}.

To help the reader understand Eq. (4), let's present an example of the calculation of p(a1), assuming the following situation:

C = {a, b, c}
$k_1 = 3$ $s_1 = 2$ and thus, $n_c = 3!/(3-2)!/2! = 3$

Then, the n_c s_o and s_a samples = {ab, ac, bc}

For simplicity, assume that each sample in C has an equal probability of being selected and thus, $p_o = p_a = 1/3$

Then using (4),

$$p(a1) = \frac{1}{2}\sum_{o=1}^{3}\sum_{a=1}^{3} s_{1o}©s_{1a}\frac{1}{3}\frac{1}{3} = \frac{1}{18}\sum_{o=1}^{3}\sum_{a=1}^{3} s_{1o}©s_{1a} \quad (5)$$

In (5) the double summation corresponds to the sum of counts of coincidences between each sample s_o and s_a, for example:

$s_{11} © s_{11}$ = ab © ab = 2
$s_{11} © s_{12}$ = ab © ac = 1 and so on until,
$s_{13} © s_{13}$ = bc © bc = 2

For this example, each term of the double summation is

$s_{1o} © s_{2a}$ = {2,1,1,1,2,1,1,1,2} and hence p(a1) = 12/18 = 2/3

Moreover, note that the calculated value for p(a1) is exactly equal to $s_1/k_1 = 2/3$. Indeed, it can be demonstrated[4] (not included here due to space limitations) that when all samples are equally probable (i.e., $p_o = p_a = 1/n_c$), which corresponds to the situation in the example, Eq. (4) reduces to

$$p(a1) = \frac{s_1}{k_1} \tag{6}$$

Furthermore, though we have no formal demonstration yet, it appears that (6) is a lower bound for p(a1). This is reasonable because if all samples (i.e., combinations of properties) were equiprobable, and we increased the probability of one of those samples, then there would be an increased probability of that sample being selected as reference s_{1o} and as comparison s_{1a} sample. By (4), this means that p(a1) increases. In contrast, if we were to decrease the probability of one of those samples, expression (2) means that other samples should increase in probability, meaning that p(a1) could not get smaller than s_1/k_1.

Regardless of whether Eq. (6) is a lower bound or not, it aids our understanding of Eq. (4). Because (6) makes p(a1) readily understandable as the number of property types coherent with a version of concept C in an average individual's mind, over the total number of property types available for C, it highlights that p(a1) is a measure of the coherence of a conceptual representation in the minds of a social group. Greater coherence implies higher p(a1).

4.4. Probability of illusory agreement, p(a2)

Recall that illusory agreement occurs when an observer O, who is in conceptual state C, witnesses A doing or uttering i, and i is consistent with O's conceptual state given that A in fact construed an alternative conceptualization Cn. In terms of the situation stipulated in Fig. 2, imagine that sample s_{1o} taken from C is chosen as the reference sample (i.e., O's sample) and then that another s_{2a} sample is taken from Cn for comparison (i.e., A's sample). Imagine also that in this second s_{2a} sample, one property i is chosen at random (i.e., whatever A decides to say or do in a specific situation). The illusory agreement probability (p(a2)) is the probability that this i property is also contained in the s_{1o} reference sample (i.e., that O sees it as evidence in favor of her own conceptualization).

To compute p(a2) we must first calculate the probability that a sample taken from Cn offers evidence consistent with O's concept. Because A will draw a sample of size s_2 properties, the sample will offer consistent evidence to O, if a property in A's sample coincides with one of the properties of the sample drawn by O from C. Thus, each sample drawn by A will have a probability of providing positive evidence, which will be directly proportional to the number of common properties between A's and O's samples. Assuming that each property in A's sample has an equal probability of being offered by A, then the probability is simply the number of common properties between A's and O's samples, divided by s_2. Because we will have many possible samples drawn from C and Cn, and each will have a different probability of being drawn, we must take an expectation of the individual probabilities of offering consistent evidence for calculating p(a2).

Recall that to formulate p(a1), we needed to compute the number n_c of possible samples taken from C. Similarly, to formulate p(a2), we need to compute the number n_{cn} of possible samples taken from Cn.

$$n_{cn} = \frac{k_2!}{(k_2 - s_2)!s_2!} \tag{7}$$

Clearly, these n_{cn} samples will have their own probability distribution, so

$$\sum_{i=1}^{n_{cn}} p(s_{1i}) = 1 \tag{8}$$

Then, just as done for p(a1), we can formulate p(a2) as

$$p(a2) = E\left[\frac{s_{11}©s_{21}}{s_2} + \frac{s_{11}©s_{22}}{s_2} + \ldots + \frac{s_{1nc}©s_{2ncn}}{s_2}\right] \tag{9}$$

Recalling that we are taking expectation over the probability of being drawn for each of the n_c samples taken from C and the same probability for each of the n_{cn} samples taken from Cn, and assuming that the drawings are independent from each other, we can rewrite Eq. (9) as:

$$p(a2) = \frac{1}{s_2}\sum_{o=1}^{nc}\sum_{a=1}^{ncn} s_{1o}©s_{2a}p_o p_a \tag{10}$$

We could further develop expression (10) to more formally define the coincidence operator © using indicator variables, but we think that would not add any value to the present paper. Instead, we think that to aid the reader in understanding Eq. (10), it will be useful to show an example of the calculation of p(a2).

Let's use the same example as before, but adding the necessary information regarding Cn:

$C = \{a, b, c\}$ $Cn = \{b, c, d, e\}$
$k_1 = 3$ $s_1 = 2$ $k_2 = 4$ $s_2 = 3$
$u = 2$ (two common elements, i.e. b,c) and thus,
$n_c = 3!/(3-2)!/2! = 3$ $n_{cn} = 4!/(4-3)!/3! = 4$
the n_c s_o samples = {ab, ac, bc} the n_{cn} s_a samples = {bcd, bce, bde, cde}

For simplicity, assume that each sample in C and in Cn has an equal probability of being selected and thus,

$p_o = 1/3$
$p_a = 1/4$

Then,

$$p(a2) = \frac{1}{3}\sum_{o=1}^{3}\sum_{a=1}^{4} s_{1o}©s_{2a}\frac{1}{3}\frac{1}{4} = \frac{1}{36}\sum_{o=1}^{3}\sum_{a=1}^{4} s_{1o}©s_{2a} \tag{11}$$

[4] Please contact the authors if interested in this (somewhat lengthy) demonstration.

In (11) the double summation corresponds to the sum of counts of coincidences between each sample s_o and s_a, for example:

$s_{11} © s_{21} = \text{ab} © \text{bcd} = 1$
$s_{11} © s_{22} = \text{ab} © \text{bce} = 1$ and so on until,
$s_{13} © s_{24} = \text{bc} © \text{cde} = 1$

For this example, each term of the double summation is

$s_{1o} © s_{2a} = \{1,1,1,0,1,1,0,1,2,2,1,1\}$ and hence $p(a2) = 12/36 = 1/3$

Note from (10) that p(a2) is influenced by all variables in the theory. For reasons that should become clear later, we want to focus here on the variables located inside the double summation in (10). It is clear that p(a2) will increase if any of the terms in the double summation increases. Hence, p(a2) will increase if the number of properties in the intersection increases, also increase if the probability of samples in C that contain properties in the intersection increases, and it will also increase if the probability of samples in Cn that contain properties in the intersection increases.

5. Applying CAT to thinking about reference to diffuse objects

To recapitulate, CAT is a theory about the conditions in which O experiences agreement with A in reference to a diffuse (and sometimes unstated) entity defined by a conceptualization. By manipulating different variables in the theory's equations, CAT is able to make predictions and give novel accounts of apparently disparate phenomena. As an illustration of this, we will briefly discuss two different phenomena and how CAT handles them. Of course, much of this discussion is at this point speculative, but we believe it succeeds in showing the theory's power and scope.

5.1. Stereotypes and the effect of illusory agreement

Stereotypes are concepts. They describe a diffuse entity (a social group) in a certain way, and explain those people's behaviors by appealing to their construed characteristics (e.g., Latinos are family oriented). In the stereotype literature, perhaps the most intriguing finding is that stereotypes (in particular, gender stereotypes) fail to fade away even when social conditions suggest they should (see Prentice & Carranza, 2004, for an in depth discussion). Gender stereotypes persist even in societies that promote an increasingly egalitarian role distribution between men and women (see Rudman & Phelan, 2008).

CAT explains this stability as follows. Consider a stereotyped pair C and Cn (e.g., *femininity* and *masculinity*), and imagine that a social group wants to reduce stereotyped judgments (e.g., "primary school teacher is a feminine profession"). At first glance, it appears intuitive that increasing the number of primary school teachers that are male (i.e., promoting a more egalitarian role or property type distribution), would reduce stereotypes. The theory behind this intuition (for a critical discussion, see Dunham, Baron, & Banaji, 2008) is that people learn and maintain stereotypes by observation of correlations between properties over relatively large learning sets (e.g., primary school teachers tend to be women, not men), and that therefore, changing the role distribution would change stereotypes. In contrast to this intuition, CAT adopts a different view by asking what would an observer O see in this new modified environment (i.e., the high male teacher proportion society), given that O has the problematic stereotype in her mind. To answer this question, recall that increasing the probability of finding tokens of property types that are in the $C \cap Cn$ set, increases p(a2) (see our discussion of Eq. (10)). In our particular example, each time that O has *femininity* in her mind, and learns of a male primary school teacher, O has an increased probability of feeling agreement, regardless of whether that particular male primary school teacher had *femininity* or not when choosing that occupation (i.e., increased p(a2)). In other words, O will have the feeling that people have become increasingly feminine, not necessarily that the concept of femininity has changed.

If stereotypes emerge and are maintained because they promote agreement of any type, and because changes in property type distribution, in particular those changes in the $C \cap Cn$ set, can increase felt agreement among a population that uses those concepts, then this would introduce strong pressures for stereotype stability. To test the internal validity of this explanation, we set up an Agent Based Model (ABM) computational simulation that implements CAT and uses agreement to increase or decrease a concept's salience in individual agents' minds. In these simulations, agents share conceptual content for two partially overlapping and contrasting C and Cn concepts (e.g., *male* and *female*), but different agents have different versions of these concepts (i.e., there are differences across individuals regarding the precise content of both *male* and of *female*). In each simulation cycle, agents observe another agent's behavior and check if their behavior is consistent with their own version of the currently focal concept (e.g., *female*). This part of the simulation models individuals who look for confirmation in a conceptual environment consistent with CAT's assumptions (though they are not biased toward confirmation; cf., Nickerson, 1998). As agents accrue confirming or disconfirming evidence, their propensity to act according to their focal concept (which we call the concept's *salience*) changes linearly (e.g., the more evidence of *female* consistent behavior in her environment, the more likely the individual is to act according to this stereotype). In the real world, this part of the simulation could correspond to associative mechanisms that influence behavior in an automatic fashion (cf., Bargh & Williams, 2006). However, because conceptual content is only partially shared, even if two agents have similar C concepts, there is only a p(a1) probability that agents receive evidence of a shared conceptualization. Also, because C and Cn overlap, there is a p(a2) probability that agents receive evidence of a shared conceptualization when agents really instantiate contrasting conceptualizations (e.g., interpreting an action as evidence of someone conceptualizing a situation as an instance of *femininity* when that person actually conceptualized it as an instance of *masculinity*).

In the simulations, our model behaved consistently with our mathematical and theoretical formulations. Increasing p($a2$) (i.e., increasing the overlapping content of the contrasting concepts), led to an increase of the stereotyped concept's salience in our agents' minds. Furthermore, sensitivity analyses and systematic testing with the model revealed that system dynamics were affected by agreement more than by other parameters of the model, consistently with our theoretical formulations. All this is not to say that stereotype change is not possible in CAT. Our simulations showed that stereotypes would tend to disappear in a social group, if the s:k ratio (that influences p($a1$)) went below a certain threshold, meaning that the number of property types available to talk about men and women was much greater than the number of properties subsumed under the stereotypes, making these concepts useless for agreement (for further detail regarding the ABM and corresponding results, we refer the interested reader to Canessa, Chaigneau, & Quezada, 2011; Chaigneau, Canessa, & Quezada, 2011).

This general idea that illusory agreement promotes stereotype stability has been previously explored in the *false consensus* literature (e.g., Bosveld, Koomen, van der Pligt, & Plaisier, 1995; Fabrigar & Krosnic, 1995; Marks & Miller, 1987; Ross, Green, & House, 1977; Strube & Rahimi, 2006). However, in this literature, false consensus has been generally defined as the (often emotional) tendency to attribute to others the views held by the self. In CAT, in contrast, it is viewed as the result of the way meaning is structured.

5.2. Public opinion

There are two ways to understand public opinion (Price, 1992). One view is that public opinion consists of informed and mostly rational stances regarding specific social issues (e.g., support for using government funds to aid banks in distress during a crisis of the economy). The other view is that public opinion is formed by mostly emotional attitudes (i.e., values) toward general classes of stimuli (e.g., adherence to a political party, support for social welfare in general). On both views, an important problem is people's lack of consistency regarding public issues (Converse, 1964). Not only are people prone to produce pseudo-opinions on issues about which they are mostly ignorant (e.g., Bishop, Oldendick, Tuchfarber, & Bennett, 1980), but they also fail to show coherence among their different opinions, and temporal stability in their stated positions (Converse, 2000; Sniderman & Bullock, 2004). However, people's opinions on public issues are not entirely random, and several decades of research focus on factors that affect stated opinions (many of which are reviewed in Bishop, 2005).

The problem of people's lack of consistency in the public opinion literature, bears a remarkable similarity to the problem of inter and intrasubjective variability that has worried theorists of concepts. Researchers on public opinion and researchers on concepts, have both worried that variability points not merely to measurement error, but to issues of validity instead. In the concepts literature, validity refers to the question of whether concepts exist as entities in the mind. In the public opinion literature, validity refers to the question of whether public opinion exists (for a discussion, see Bishop, 2005). However, as explained earlier, rather than being a problem for CAT, differences in conceptual content promote true and illusory agreement, and this mechanism may be able to account for the dynamics of public opinion, as the discussion that follows illustrates.

Studies of partisanship in different countries (as measured by opinion polls), consistently describe that political preferences at an aggregate level produce time series in which periods of relatively persistent inclinations for one political option over another, are quasi-periodically followed by reversals (e.g., Byers, Davidson, & Peel, 2000; Box-Steffensmeier & Tomlinson, 2000; DeBoef, 2000; Lebo, Walker, & Clarke, 2000; Maestas & Preuhs, 2000; Wlezien,

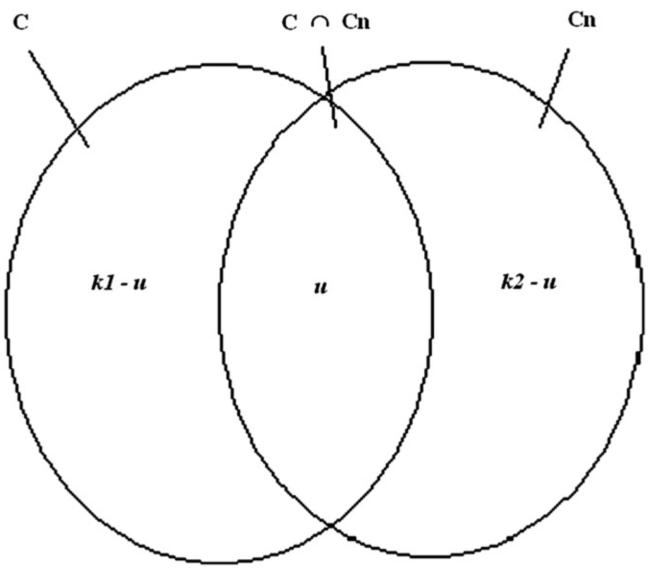

Fig. 3. Venn diagram illustrating the intersection of two concepts C and Cn, where $k1$ = number of properties in C, $k2$ = number of properties in Cn, and u = number of properties in the intersection between C and Cn.

2000). An explanation that has been provided for these dynamics is that voters vary in the strength of their party preferences, such that some are relatively insensitive to contingent information (i.e., the "committed" voters) and others are more sensitive to it (i.e., the "floating" voters). Though we oversimplify this theory by putting it on these terms, presumably the committed voters are the source of stability in the time series, and the floating voters account for reversals (Box-Steffensmeir & Smith, 1996; Byers et al., 2000).

In contrast, CAT would explain the same data, making very different assumptions. Preliminary agent-based modeling efforts in our laboratory show that quasi-periodic aggregate data can be modeled by a group of computational agents that have two partially overlapping *C* and *Cn* concepts in their minds (e.g., *left*- and *right-wing*), and where each concept can control an agent's behavior depending on its relative salience, with salience depending on a concept's ability to generate true or illusory agreement (the more agreement of any kind that the concept affords for the agent, the more salient the concept becomes and the more able to direct behavior). It turns out that if *C* and *Cn* are partially opposed and partially overlapping (as illustrated in Figs. 1–3), and thus promoting true and illusory agreement, this simple system exhibits quasi-periodical shifts on conceptual salience, with *C* and *Cn* alternating their control of agents' behavior. The reason for these dynamics is that as a concept *C* increases its salience, it incrementally controls agents' behavior but also provides increasing illusory agreement opportunities for agents that express the contrast concept *Cn*. This produces periods of dynamical stability with quasi-periodic shifts in conceptual salience. (We hasten to add that these preliminary findings are not proof of our theory's external validity).

6. Conclusion

CAT is a theory about agreement, and the relevance of agreement for our social life is emphasized by many social philosophers such as Habermas (1989) and Searle (1995). The latter, for example, points out that agreement is a condition of possibility for the whole of social life, and that social objects like money, marriage and government are not independent from human agreement. In fact, there is a wide range of situations where agreement about diffuse entities is necessary for human action, going from everyday conversation to the achievement of scientific discoveries.

To achieve coordination in many complex everyday activities, we need to believe that we refer to *the same thing*. Our roles in society, for example, presuppose multiple explicit, but above all implicit, agreements about what is expected of us as parents, spouses, or members of an organization. Similarly, our emotional relationships often depend on agreeing that we have some real and stable personal characteristics, and that these may be known.

In less common situations, people sometimes develop agreements about things that are perhaps even more diffuse. Some clinical psychologists, for example, talk about *ego structure* and believe that a fragmented ego structure relates to psychosis, a disordered one relates to borderline personality structure, and that a rigid ego structure relates to neurosis (e.g., Kernberg & Caligor, 2005), in which case they may advice psychotherapy, which is a conversation where patient and psychotherapist talk about who the patient really is, until both can agree that the patient has changed and treatment can be suspended.

In scientific conversations, scientists need to agree on diffuse entities such as what is an interesting problem for research, what counts as a solution, and even what is it that they do research on. Neuroscientific research on brain functions, for example, requires such agreements. As Bechtel and McCauley (1999) discuss, if these agreements were not achieved, comparative neuroscience would be practically impossible.

What we try to highlight with these examples, is that diffuse referents are everywhere to be found. Many signs and symbols point to them (e.g., a flag, a city landmark, a wedding ring), and our ability to agree that these symbols point to the same thing is crucial to our lives. We believe CAT can help us comprehend how these agreements in individual understandings develop in human groups, and what the consequences of those agreements are for the fate of the concepts that promote them.

References

Antaki, C. (2007). Mental health practitioner's use of idiomatic expressions in summarizing clients' accounts. *Journal of Pragmatics, 39*, 527–541.

Arístegui, R., Gaete, J., Muñoz, G., Salazar, J. I., Vilches, O., Krause, M., et al. (2009). Diálogos y autorreferencia: procesos de cambio en psicoterapia desde la perspectiva de los actos de habla. [Dialogues and self-reference: change in psychotherapy from the perspective of speech acts]. *Revista Latinoamericana de Psicología, 41*(2), 277–289.

Atkinson, J. M., & Heritage, J. (Eds.). (1984). *Structures of social action: Studies in conversation analysis*. Cambridge, MA: Cambridge University Press.

Bargh, J. A., & Williams, E. L. (2006). The automaticity of social life. *Current Directions in Psychological Science, 15*, 1–4.

Barsalou, L. W. (1987). The instability of graded structure: implications for the nature of concepts. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors is categorization* (pp. 101–140). Cambridge: Cambridge University Press.

Barsalou, L. W. (1993). Flexibility, structure, and linguistic vagary in concepts: manifestations of a compositional system of perceptual symbols. In A. C. Collins, S. E. Gathercole, & M. A. Conway (Eds.), *Theories of memory* (pp. 29–101). London: Lawrence Erlbaum Associates.

Bechtel, W., & McCauley, R. N. (1999). Heuristic identity theory (or back to the future): the mind-body problem against the background of research strategies in cognitive neuroscience. In M. Hahn, & S. C. Stoness (Eds.), *Proceedings of the 21st Annual Meeting of the Cognitive Science Society* (pp. 67–72). Mahwah, NJ: Lawrence Erlbaum Associates.

Berlin, B. (1992). *Ethnobiological classification: Principles of categorization of plants and animals in traditional societies*. Princeton, NJ: Princeton University Press.

Billig, M. (1996). *Arguing and thinking: A rethorical approach to social psychology*. Cambridge: Cambridge University Press.

Bishop, G. F. (2005). *The illusion of public opinion: Fact and artifact in American public opinion polls*. Lanham, MD: Rowman & Littlefield Publishers, Inc.

Bishop, G. F., Oldendick, R. W., Tuchfarber, A. J., & Bennett, S. E. (1980). Pseudo-opinions on public affairs. *Public Opinion Quarterly, 44*(2), 198–209.

Bosveld, W., Koomen, W., van der Pligt, J., & Plaisier, J. W. (1995). Differential construal as an explanation of false consensus and false uniqueness effects. *Journal of Experimental Social Psychology, 31*, 518–532.

Box-Steffensmeier, J. M., & Tomlinson, A. R. (2000). Fractional integration methods in political science. *Electoral Studies, 19*(1), 63–76.

Box-Steffensmeir, J. M., & Smith, R. M. (1996). The dynamics of aggregate of partisanship. *American Political Science Review, 90*, 567–580.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1482–1493.
Brown-Schmidt, S., & Tanenhaus, M. K. (2008). Real-time investigation of referential domains in unscripted conversation: a targeted language game approach. *Cognitive Science, 32*, 643–684.
Byers, D., Davidson, J., & Peel, D. A. (2000). The dynamics of aggregate political popularity: evidence from eight countries. *Electoral Studies, 19*, 49–62.
Canessa, E., Chaigneau, S. E., & Quezada, A. (2011). An ABM of the development of shared meaning in a social group. In J. Filipe, & A. Fred (Eds.), *Proceedings of the 3rd international conference on agents and artificial intelligence, Vol. 2* (pp. 5–14).
Carpenter, M., Nagell, K., & Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development, 63*(4), 1–143.
Chaigneau, S. E., Canessa, E., & Quezada, A. (2011). The spreading and demise of concepts in social groups. In F. Ochoa, F. Meza, D. Mery, & C. Cubillos (Eds.), *IEEE Proceedings of the XXIX International Conference of the Chilean Computer Science Society (SCCC 2010)* (pp. 139–145).
Channell, J. (1994). *Vague language*. Oxford: Oxford University Press.
Clark, H. H., & Krych, M. A. (2004). Speaking while monitoring addressees for understanding. *Journal of Memory and Language, 50*(1), 62–81.
Converse, P. E. (1964). The nature of belief systems in mass publics. In D. E. Apter (Ed.), *Ideology and discontent* (pp. 206–261). New York: The Free Press.
Converse, P. E. (2000). Assessing the capacity of mass electorates. *Annual Review of Political Science, 3*, 331–353.
Davis, K. (1986). The process of problem (re)formulation in psychotherapy. *Sociology of Health & Illness, 8*(1), 44–47.
DeBoef, S. (2000). Persistence and aggregations of survey data over time: from microfoundations to macropersistence. *Electoral Studies, 19*(1), 9–29.
Dunham, Y., Baron, A. S., & Banaji, M. R. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences, 12*(7), 248–253.
Edwards, D. (1997). *Discourse and cognition*. London: Sage.
Fabrigar, L. R., & Krosnic, J. A. (1995). Attitude importance and the false consensus effect. *Personality and Social Psychology Bulletin, 21*, 468–479.
Frege, G. (1893/1952). On sense and reference. In P. Geach, & M. Black (Eds.), *Translations from the philosophical writings of Gottlob Frege* (pp. 56–78). Oxford: Blackwell.
Galantucci, B., & Sebanz, N. (2009). Joint action: current perspectives. *Topics in Cognitive Science, 1*, 255–259.
Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
Garrod, S., & Anderson, A. (1987). Saying what you mean in dialogue: a study in conceptual co-ordination. *Cognition, 27*, 181–218.
Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy? *Trends in Cognitive Sciences, 1*(2), 292–304.
Gergen, K. J. (1991). *The saturated self: Dilemmas of identity in contemporary life*. London: Harper.
Glock, H. J. (2009). Concepts: where subjectivism goes wrong. *Philosophy, 84*(1), 5–29.
Habermas, J. (1989). *Teoría de la acción comunicativa I: Racionalidad de la acción y racionalización social*. [*The theory of communicative action: Reason and the Rationalization of Society*]. Buenos Aires, Argentina: Alfaguara.
Hampton, J. A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18*, 441–461.
Horwich, P. (1998). *Meaning*. Oxford: Oxford University Press.
Jucker, A., Smith, S., & Lüdge, T. (2003). Interactive aspects of vagueness in conversation. *Journal of Pragmatics, 35*, 1737–1769.
Kernberg, O., & Caligor, E. (2005). A psychoanalytic theory of personality disorders. In M. F. Lenzenweger, & J. F. Clarkin (Eds.), *Major theories of personality disorder* (2nd ed.). (pp. 114–156) New York, NY: The Guilford Press.
Kripke, S. (1980). *Naming and necessity*. Cambridge, MA: Harvard University Press.
Lebo, M. J., Walker, R. W., & Clarke, H. D. (2000). You must remember this: dealing with long memory in political analyses. *Electoral Studies, 19*(1), 31–48.
Leudar, I., Sharrock, W., Truckle, S., Colombino, T., Hayes, J., & Booth, K. (2008). Conversation of emotions: on turning play into psychoanalytic psychotherapy. In A. Peräkylä, C. Antaki, S. Vehviläinen, & I. Leudar (Eds.), *Conversation analysis and psychotherapy* (pp. 152–172). Cambridge, MA: Cambridge University Press.
Levinson, S. (1983). *Pragmatics*. Cambridge, MA: Cambridge University Press.
Lewis, D. (1969). *Convention: A philosophical study*. Cambridge: Harvard University Press.
Lewis, D. (1975). Languages and language. In K. Gunderson (Ed.), *Minnesota studies in the philosophy of science, Vol. VII* (pp. 3–35). Minneapolis: University of Minnesota Press.
Maestas, C., & Preuhs, R. R. (2000). Modeling volatility in political time series. *Electoral Studies, 19*(1), 95–110.
Marks, G., & Miller, N. (1987). Ten years of research on the false-consensus effect: an empirical and theoretical review. *Psychological Bulletin, 102*, 72–90.
Marsh, K. L., Richardson, M. J., & Schmidt, R. C. (2009). Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science, 1*, 320–339.
McHoul, A. W., & Rapley, M. (2001). *How to analyse talk in institutional settings: A casebook of methods*. New York, NY: Continuum.
McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavioral Research Methods, Instruments, and Computers, 37*, 547–559.
Medin, D. L., Altom, M. W., Edelson, S. M., & Freko, D. (1982). Correlated symptoms and simulated medical classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 37–50.
Millikan, R. G. (2005). *Language: A biological model*. Oxford: Oxford University Press.
Moses, L. J., Baldwin, D. A., Rosicky, J. G., & Tidball, G. (2001). Evidence of referential understanding in the emotions domain at 12 and 18 months. *Child Development, 72*, 718–735.
Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
Murphy, G. L., & Wisniewski, E. J. (1989). Feature correlations in conceptual representations. In G. Tiberghien (Ed.), *Theories and application. Advances in cognitive science, Vol. 2* (pp. 23–45). Chichester, England: Ellis Horwood.
Nickerson, R. (1998). Confirmation bias: a ubiquitous phenomenon in many guises. *Review of General Psychology, 2*, 175–220.
Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115*, 1939–1957.
Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(1), 87–108.
Prentice, D. A., & Carranza, E. (2004). Sustaining cultural beliefs in the face of their violation: the case of gender stereotypes. In M. Schaller, & C. S. Crandall (Eds.), *The psychological foundations of culture* (pp. 259–280). Mahwah, NJ: Lawrence Erlbaum Associates.
Price, V. (1992). *Public opinion*. Newbury Park, CA: Sage Publications.
Putnam, H. (1973). Meaning and reference. *The Journal of Philosophy, 70*(19), 699–711.
Quine, W. V. O. (1951). Two dogmas of empiricism. *The Philosophical Review, 60*, 20–43.
Quine, W. V. O. (1969). *Ontological relativity and other essays*. New York, NY: Columbia University Press.
Richardson, D. C., Dale, R., & Tomlinson, J. M. (2009). Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science, 33*, 1468–1482.
Rosch, E., & Mervis, C. B. (1975). Family resemblances: studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573–605.
Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.
Ross, L., Green, D., & House, P. (1977). The false consensus effect: an egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology, 13*, 279–301.
Rudman, L. A., & Phelan, J. E. (2008). Backlash effects for disconfirming gender stereotypes in organizations. *Research in Organizational Behavior, 28*, 61–79.
Russell, B. (1997). *The problems of philosophy*. Oxford: Oxford University Press.
Sacks, H. (1972). An initial investigation of the usability of conversational data for doing sociology. In D. Sudnow (Ed.), *Studies in social interaction* (pp. 31–74). New York, NY: The Free Press.
Santos, A., Chaigneau, S. E., Simmons, W. K., & Barsalou, L. W. (2011). Property generation reflects word association and situated simulation. *Language and Cognition, 3*(1).
Searle, J. R. (1995). *The construction of social reality*. New York, NY: The Free Press.
Shoter, J. (1993). *Conversational realities: Constructing life through language*. London: Sage.

Simmons, W. K., Hamann, S. B., Harenski, C. N., Hu, X. P., & Barsalou, L. W. (2008). fMRI evidence for word association and situated simulation in conceptual processing. *Journal of Physiology – Paris, 102*, 106–119.
Sniderman, P. M., & Bullock, J. G. (2004). A consistency theory of public opinion and political choice: the hypothesis of menu dependence. In W. E. Saris, & P. M. Sniderman (Eds.), *Studies in public opinion: Gauging attitudes, nonattitudes, measurement error, and change* (pp. 337–357). Princeton, NJ: Princeton Univesity Press.
Sperberg, D., & Wilson, D. (1991). Loose talk. In S. Davis (Ed.), *Pragmatics: A reader* (pp. 540–549). New York, NY: Oxford University Press.
Spivey, M. J., Tanenhaus, M. K., Eberhard, K. M., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45*(4), 447–481.
Strong, T., Busch, R., & Couture, S. (2008). Conversational evidence in therapeutic dialogue. *Journal of Marital and Family Therapy, 34*, 388–405.
Strube, M. J., & Rahimi, A. M. (2006). "Everybody knows it's true": social dominance orientation and right-wing authoritarianims moderate false consensus for stereotypic beliefs. *Journal of Research in Personality, 40*, 1038–1053.
Tomasello, M. (1995). Joint attention as social cognition. In C. Moore, & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Hillsdale, NJ: Lawrence Erlbaum.
Tversky, A. (1977). Features of similarity. *Psychological Review, 84*(4), 327–352.
Van der Howen, F. (2009). Formulating disputes. *Journal of Pragmatics, 41*, 2072–2085.
Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: an updated and expanded version of the norms. *Journal of Memory and Language, 50*(3), 289–335.
Wlezien, C. (2000). An essay on 'combined' time series processes. *Electoral Studies, 19*(1), 77–93.