Concepts as Representations for Essences: Evidence from Use of Generics

Iris Oved[1] (irisoved@gmail.com)
Pierina Cheung[2] (mpcheung@uwaterloo.ca)
David Barner[1] (barner@ucsd.edu)

[1] Department of Psychology, University of California, San Diego 9500 Gilman Drive, San Diego, CA 92093, USA

[2] Department of Psychology, University of Waterloo 200 University Avenue W., Waterloo, Ontario, N2L 3G1, Canada

Abstract

This paper compares descriptivist approaches for concept acquisition with essentialist approaches by exploring the conditions under which people use generic sentences (sentences such as 'Apples are round' which contrast with sentences about particulars like 'All/most of the apples are round'). It fleshes out the essentialist approach in terms of the Baptism theory of concept acquisition (Oved, 2009; 2014), which is made precise with an implementation in which concepts are values of latent variables in a Bayesian network, posited as explanations for observed patterns in objects' perceptible properties (Oved & Fasel, 2011). Two experiments measuring the use of generics are described and used as support for this essentialist approach over descriptivist approaches.

Keywords: Concepts; Concept-Acquisition; Essences; Explanation; Bayesian Networks; Generics.

Introduction

When we hear, "these are kiwanos", while being shown a set of objects each displaying a novel combination of yellowness, spikiness, and ovalness, how do these regularities factor into our hypotheses of what 'kiwano' means? One possibility is that we acquire a new concept, KIWANO, by conjoining the representations YELLOW, SPIKEY, and OVAL such that the observed regularities are directly encoded as part of the meaning of the concept. Perhaps over time, and over multiple learning instances, criterial information is differentiated from noise, and we come to represent KIWANO as a set of probability distributions over the most typical features. Variants of this *descriptivist approach* include classical definition theories (Locke, 1690; Hume, 1748), prototype theories (Rosch, 1978; Barsalou, 1987, 1999; Prinz 2002), and exemplar theories (Smith & Medin 1981). Another possibility, however, is that observed regularities are not themselves constitutive of meanings, but instead serve only as the basis for positing hidden/latent properties to explain the regularities. On such *essentialist accounts*, the learned concept KIWANO may be simple in representational structure (Gelman, 2003; Keil, 1989; Carey, 1985).

The present paper articulates a favored version of the essentialist approach (Oved, 2009; 2014; Oved & Fasel 2011), an account in the spirit of Bayesian cognitive science (following Griffiths, Kemp, & Tenenbaum, 2008; Xu & Tenenbaum, 2007; Gopnik & Tenenbaum, 2007; Kemp, Perfors, & Tenenbaum, 2007). The paper then describes two experiments on the conditions for use of generic sentences that support the essentialist approach over the descriptivist approach.

Baptizing Concepts for Hypothesized Kinds

Descriptivist accounts of concepts are largely motivated by the idea that many of our concepts are acquired as the result of observed associations, and the most straight-forward way to acquire them is to compose them directly from perceptual representations. Nativists since Plato have insisted, however, that many of our concepts cannot be composed from (even probability distributions over) perceptual representations (for arguments see Fodor, 1975, 1981, 1998, 2008). The Baptism view of concept acquisition is motivated by *Fodor's Challenge* – the challenge of showing how a representationally simple concept can be learned from observed associations, *genuinely* learned in the sense of hypothesis formation and testing (Oved, 2009; 2014).

The view is inspired by Kripke's (1972) baptism view on the analogous question of *how proper names in natural languages come to have their meanings*, treating the concept-acquisition question as the question of *how mental names for properties/kinds come to have their meanings*. Kripke's maneuver for the meanings of proper names was to use a reference-fixing description that involves deictic representations, or pointers – e.g., '*this* baby in *my* arms *now*', to avoid identifying proper names with descriptions. A similar approach is used in the Baptism account for concepts, where the reference-fixing description, e.g., for the concept APPLE, might be, 'the latent/unobservable property that *these* objects have that explains *this* similarity (in redness, roundness, sweetness, etc.)'. Notice that the description appeals to deep/hidden/latent properties, which is what makes the account a species of psychological essentialism. As long as this description manages to uniquely pick out the property of appleness, it can serve as a step in acquiring a simple mental term for appleness.[1]

To see that the Baptism process answers Fodor's challenge of showing how a representationally simple

[1]Note that a different description could have been used, say, using taste and touch information, perhaps by a blind child, *so long as it too manages to pick out appleness.* See (Oved, 2009; 2014) for details on how to deal with what philosophers have called the qua-problem (Devitt & Sterelny, 1999). Roughly, different sets of similarities are explained by appleness than by fruitness, edibleness, organicness, McIntoshness, etc..

concept can be learned by hypothesis formation and testing, we must step back to before the concept learner forms the reference-fixing description. Consider the process in steps.

Step 1: Assume deep structure. As a species of psychological essentialism, the Baptism theory claims that humans assume by default that the many similarities and differences in objects' observable properties are determined and explained by a smaller set of latent/unobservable properties. These latent properties may be understood as *essences*, bearing in mind that any given object may have several latent properties that each explain different sets of observable properties. One way to implement this assumption is with a Bayesian network like that depicted in Figure 1, which was used in a software robot baby that learned the number of fruit-kinds in its world as well as the dependencies of the objects' colors and shapes on their fruit kinds (Oved & Fasel, 2011). In contrast, a descriptivist model would only store shallow associations between observable classes of properties (e.g., between colors and shapes), leaving out the further class of properties, fruit-kinds, represented in this model. Fruit-kinds would instead be identified with probabilistic associations between observable properties.

Crucially, assuming that objects have latent classes properties that determine/explain the objects' observable properties is not equivalent to assuming that concepts have particular values of those latent classes of properties (i.e., it does not assume Fodor's radical nativism about simple concepts). Particular values for the fruit-kinds, such as *appleness*, are not represented until experience gives reason to posit them as explanations for observed regularities.

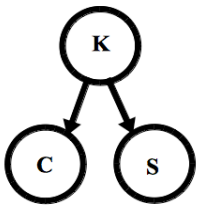

Figure 1: Bayesian net in which an object's fruit-kind, *K,* determines the object's color, *C*, and shape, *S*.

Step 2: Observation. In a world that has regularities, like ours, a concept learner that goes into its environment and makes sensory observations will find that objects cluster in their observable properties.

Step 3: Abduction. Next, according to the Baptism account, the concept learner makes an inference from the observed clusters to the explanation that objects in a cluster share a latent/unobservable property that determines their observed properties. This inference to the best explanation can be implemented with a truncated version of Bayes' rule.

$P(\text{Model}|\text{Data}) \propto P(\text{Data}|\text{Model}) \cdot P(\text{Model})$

A given Model here would be a network like that depicted in Figure 1, but fleshed out with a number n of kinds, $K_{1...n}$, and their probabilistic relationships to the various colors, C, and shapes S. The rule is then read as: the posterior probability that a given set of observed Data is explained by a given Model *is proportional to* the likelihood that the Data would be observed given that the Model brought it about, multiplied by the prior probability on the Model. Candidate Models, each with a different number of kinds and relationships to colors and shapes, can thus be compared to find the candidate with the highest probability.

Step 4: Naming. Finally, the agent assigns an arbitrary simple mental symbol (or *name*) to the newly hypothesized properties/kinds that it takes to be part of the best explanation for its set of observations.

In the following sections, we describe two experiments that aim to distinguish between the descriptivist and essentialist accounts of concepts by examining the use of generic utterances to describe regularities.

Generics, Laws, and Essences

By many accounts, generic statements (e.g., 'Apples are round' in contrast with 'All/most apples are round') are a linguistic tool for expressing essential relationships between kinds and their properties (see, e.g., Carlson, 1977; Prasada, 2000; Prasada & Dillingham, 2006, 2009; Prasada, Khemlani, Leslie & Glucksberg, 2013; Gelman, 2003; Gelman & Bloom, 2007). Prasada and colleagues have argued for three types of connections that are expressed with generic statements – principled connections, statistical connections, and causal connections (Prasada & Dillingham, 2006, 2009; Prasada et al., 2013). Of interest here are principled connections – properties that members of a kind have *by virtue of* being that kind of thing. For example, being round is a property that apples have by virtue of being an apple. We exploited this feature of generic statements to investigate how observable regularities are involved in concept acquisition. In two experiments, we manipulated whether a set of regularities was presented as accidental and measured the proportion of generic utterances used to describe the regularities.

Experiment 1

In the first experiment, participants were initially introduced to three different novel kinds of creatures, each colored grey. Later they were shown a world in which all members of each kind were the same color, but each kind had a distinct color. In one condition (the Default condition), the world with its regularities were simply shown without a story about how the regularities came about, allowing participants to form default interpretations of the regularities. In another condition (the Random condition), participants were shown that the regularities happened by accident. Participants in both conditions were asked to

describe the colors of the novel creatures, and we measured the proportion of sentences that were generic.

If participants store concepts as shallow associations, as suggested by descriptivist accounts, then no difference should exist between these two cases. If, however, participants by default adopt an explanatory approach to observable regularities – i.e., they take the regularities as evidence for a further property that serves as the common cause/explanation of the observed correlated traits – then a difference should exist in the proportion of generic sentences between the two conditions. Specifically, participants who were simply shown the regularity should have used relatively more generic sentences to describe the scene whereas participants who were shown that the regularity was accidental should have used relatively few generic sentences.

Participants

106 participants (mean age 34 years) were recruited via Amazon Mechanical Turk. All were located in the USA and were self-reported native English speakers. Participants were randomly assigned to two conditions: Random condition (n = 57) and Default condition (n = 49).

Stimuli

We constructed three distinct creatures each with a novel label (see Figure 3). These creatures were designed to look like animate objects. We also constructed distinct planets and labeled them each with a novel noun.

Procedure

At the beginning of the experiment, participants were told to read the slides of a story and that they would be tested on their comprehension of the story afterwards. In both the Random and Default conditions, participants first read about a distant galaxy called Plentia with planets that have three different kinds of creatures: toma, pimwit, kirbo. All creatures were presented in grey, and the names were provided on the screen.

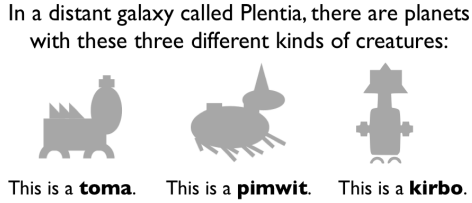

Figure 2: Initial presentation of three novel creatures shown in both Accidental and Default conditions.

In the Default condition, participants then saw the slide shown in Figure 2, with a picture of planet Gelkon with the three creatures each having a distinct color (all of the tomas were black, all of the pimwits were blue, all of the kirbos were red). Participants were reminded of the names of the creatures, and were asked to type three sentences describing the colors of the creatures on Gelkon.

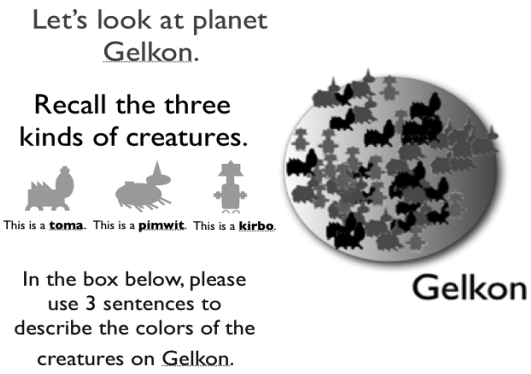

Figure 3: Final slide with probe shown to participants in both conditions.

Between the slides shown in Figures 2 and 3, participants in the Random condition read a story about how the creatures got to Gelkon. They read that at the beginning of the universe, all of the creatures were in a jar in the center of the galaxy. One day, the jar exploded and all the creatures randomly landed on the planets. Then, a series of six slides was shown to depict the explosion, with a jar containing many instances of each type of creature, uniformly distributed in green, blue, black, and red, exploding such that the animals randomly landed onto the five planets. After seeing the explosion participants in the Random condition were given the probe slide shown in Figure 3. The participants were given comprehension tests to ensure they were attending to the story.

Results

Responses were coded into three sentence types: *generic*, *particular*, and *other*. A sentence was coded as *generic* if it expressed information about the kind (e.g., 'Kirbos are red'; 'The kirbo is red'). A sentence was coded as *particular* if it referred to the particular instances (e.g., 'All of the kirbos are red'; 'The kirbos are red'). Everything else that was not generic or particular was coded as *other* (e.g., 'red, blue, black'; 'Gelkon is a populated planet'). The experimenter who coded the data was blinded to the condition the participant was in.

We computed the percentage of *generic*, *particular*, and *other* sentence types for each participant. On average across both conditions, there were 37.1% of *generic* sentences, 47.8% *particular* sentences, and 15.1% coded as *other*. Participants were highly consistent with respect to the type of response they provided: 103/106 participants always responded with the same type of sentence.

To compare how often participants in the Random and Default conditions responded with generic sentences, we conducted a Mann-Whitney U test. We found a significant difference of condition, $p = .009$, with 26.3% of *generic* responses in the Random condition and 51.0% in the Default condition (Figure 5). This suggests that participants in the Default condition were significantly more likely to describe the color of the creatures with a generic sentence.

Next, we examined whether participants in the Default condition would be more likely to respond with a generic statement than with sentences that refer to particular instances. Although the percentages suggest that participants were more likely to provide a generic response (51.0%) than a particular response (38.8%), the difference was not significant, p =.37 (Figure 4).

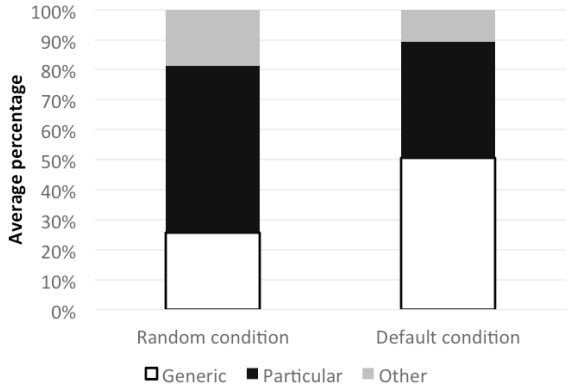

Figure 4: Percentages of Generic, Particular, and Other sentences in the two conditions.

Discussion

This experiment explored the conditions under which participants would interpret observable regularities as reflecting deep properties of a kind. Our findings suggest that, contrary to predictions from descriptivist accounts, participants were more likely to respond with generic statements when shown that the regularities were present by default (Default condition) than when shown that the regularities were accidental (Random condition). This finding suggests that participants in the Default condition were adopting an explanatory approach to observable regularities, not merely extracting shallow associations. This outcome supports the essentialist framework, in that it suggests that (a) humans distinguish between lawlike and accidental regularities when interpreting their observations and (b) by default humans infer from an observed regularity to an explanation in terms of essences.

However, the current experiment leaves open three alternative interpretations. First, participants in the Random condition *saw* instances of the kinds of creatures in different colors, and thus, they may have learned that color is not a very useful cue for category membership. Second, planets other than Gelkon were shown in the Random condition, whereas Gelkon is the only planet in the Default condition. Participants in the Random condition could have represented the creatures on Gelkon as a subset of the instances (e.g., a sample), and those in the Default condition could have represented them as the full set (e.g., the population). This set/subset difference may have led to different inferences from the observed regularities. Third, although we found a higher rate of generic responses in the Default than in the Random condition, it is not entirely clear if generic sentences truly reflect essentialist beliefs. We address each of these three concerns in Experiment 2.

Experiment 2

The goal of Experiment 2 was to replicate findings from Experiment 1 using a modified paradigm. In Experiment 2, we equated the number of instances of the creatures and also the number of planets for both the Random and Default conditions. Specifically, in both conditions, information about the relationships between properties and the kind was presented in text, and participants only saw Gelkon. Moreover, to investigate whether generic sentences reflect essentialist beliefs, we included an additional condition – Essential condition – such that the property of interest is made explicit that it is an essential property for the kind.

Participants

73 participants were recruited via Amazon Mechanical Turk. All were self-reported native English speakers who lived in the US. Participants were randomly assigned to one of three conditions: Essential condition (n = 27), Default condition (n = 23), and Random condition (n=23).

Stimuli

We used the same three creatures as Experiment 1, but their initial presentation was not filled in (see Figure 5).

Procedure

As with Experiment 1, at the beginning of the experiment, participants were told to read the slides with a story and that they would be tested on their comprehension of the story afterwards. In all of the conditions participants were told about a planet called Gelkon and were introduced to three different kinds of creatures: toma, pimwit, kirbo. All creatures were presented without any color, and the names were provided on the screen.

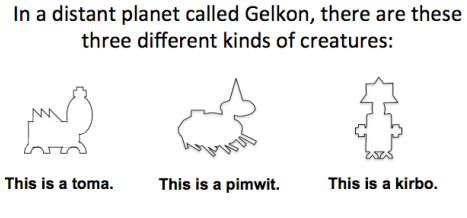

Figure 5: Initial presentation of three novel creatures shown in all conditions.

Participants in the Essential and Random conditions read a story, displayed at the bottom of the slide shown in Figure 5, about how the creatures on Gelkon got their colors. Participants in the Essential condition received a story stating that, "On Gelkon, there are different types of blood that are different colors. They type of blood a creature has determines the fur-color of the creature. For example, if a creature has green blood, then the creature will have green fur. Since a creature's blood-type stays the same for its

whole life, a creature's fur-color stays the same for its whole life." Participants in the Random condition received a story stating that the fur colors of the creatures were highly unstable and accidental. They read that, "On Gelkon, there are different types of pools that are different colors. The type of pool a creature bathes in determines the fur-color of the creature. For example, if a creature bathes in a green pool, then its fur will become green. Since a creature can bathe in a different colored pool each day, its fur-color can be different each day." Comprehension checks were made to ensure participants were reading the slides and attending to the task.

Participants in all three conditions then saw the slide shown in Figure 6, with a picture of planet Gelkon with the three creatures each having a distinct color (all of the tomas were black, all of the pimwits were blue, all of the kirbos were red). Participants were reminded of the names of the creatures, and were asked to type three sentences describing the colors of the creatures on Gelkon.

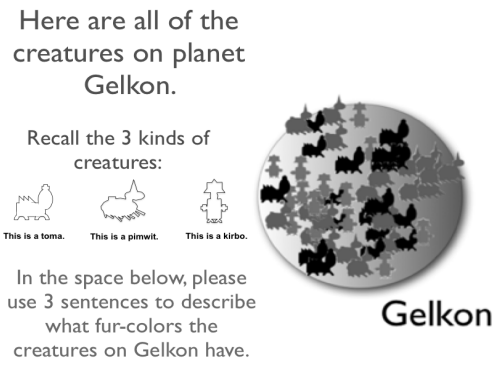

Figure 6: Final slide with probe shown to participants in all conditions.

Results

As with Experiment 1, responses were coded into three sentence types: *generic*, *particular*, and *other*. The coding was done blindly.

We computed the percentage of generic, particular, and other sentence types for each participant. On average across three conditions, there were 60.0% of 'generic' sentences, 17.5% 'particular' sentences, and 30.5% coded as 'other'.

To compare how often participants in the three conditions responded with generic sentences, we conducted a Kruskal-Wallis test and found a significant effect of condition ($H(2)=19.98$, $p < .001$). Post-hoc comparisons using Mann-Whitney U revealed a significant difference between the Random condition and both the Default ($U = 175.0$, $Z = -3.03$, $p < .001$) and Essential conditions ($U = 86.5$, $Z = -4.38$, $p < .001$). There was no significant difference between the Default and the Essential conditions ($U = 232.5$, $Z = -1.76$, $p = .078$). Figure 7 displays the proportion of the sentence types for each of the conditions. This suggests that participants in both the Essential and Default conditions were significantly more likely to describe the color of the creatures with a generic sentence.

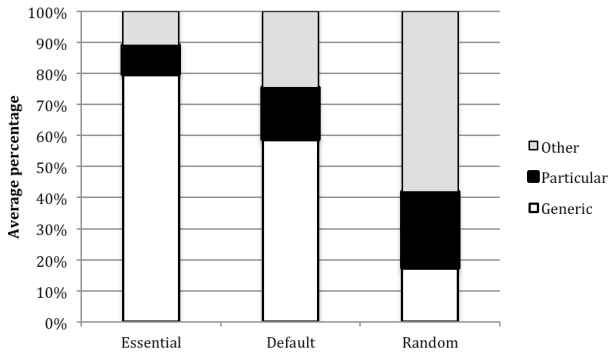

Figure 7: Percentages of Generic, Particular, and Other sentences in the three conditions.

Discussion

In Experiment 2, we found that participants in both the Essential and Default conditions were more likely to describe the color of the creatures with a generic sentence than those in the Random condition. This suggests that our results from Experiment 1 were not due to idiosyncrasies of the task. Crucially, participants from all three conditions saw the same creatures and distribution of properties. The only difference was in the information about how the property is related to the kind, which was presented in text. Participants in the Random condition read that the property of interest (i.e., color) was accidental, and those in the Essential condition read that the color was an essential property of the creatures. Moreover, our findings also suggest that generic sentences reflect essentialist beliefs.

General Discussion and Conclusions

This paper articulates an essentialist account of concepts and reports two experiments on the uses of generic sentences that supports such essentialist approaches over descriptivist ones. In Experiment 1, we aimed to investigate whether people made different inferences about how a property was related to the kind when observable regularities were accidental versus when they were lawlike. Specifically, we showed participants three novel creatures, and provided them with information about how the creatures obtained their colors. We measured the type of sentences they used to describe the colors of the creatures. Drawing on previous research on generics, we hypothesized that participants who inferred that the property-kind relation is lawlike would be more likely to describe the creatures with generic sentences than those who inferred that the property-kind relation was accidental. Results from Experiment 1 supported our hypothesis, suggesting that participants interpreted observable regularities as positing deep/hidden properties. We also found that this assumption about observable regularities is defeated when the property is believed to be related to the kind by accident. Nevertheless, the findings of Experiment 1 may also be attributable to the fact that participants in the Random condition saw more

planets (or possible worlds) and more instances that had different colors, and these differences may allow for different inferences to be made. Experiment 2 addressed these concerns by presenting information in text. Moreover, the logic of Experiment 1 rests on the assumption that generic sentences reflect a lawlike relation between the property and the kind, which was not explicitly tested. We addressed this concern in Experiment 2 by adding an Essential condition in which it was made clear that colors were an essential property of the creatures. We found that participants in both the Essential and the Default conditions were more likely to provide generic responses than those in the Random conditions.

Given that the distribution of properties was the same across the three conditions in Experiment 2, the descriptivist approach to concepts cannot explain why there was a difference in the usage of generic sentences between the Random condition and the Default and Essential conditions. Our result suggests that people interpret observable regularities as a signal for a lawlike relation between the property (color) and kinds, and we argue that this provides strong evidence for the essentialist approach to concepts.

An issue of interest in future work is how hypotheses about hidden/latent properties are revised in light of new data. Clusters of perceptually similar objects may split or merge. In some cases of such merging and splitting, the original reference-fixing descriptions may have failed either to pick out a property that exists or failed to pick out a property that is unique. Difficult puzzles thus arise about what the original concepts meant in cases of such revisions. These concerns are discussed in the full proposal of the Baptism theory of concepts (Oved, 2009, in press) but are in need of empirical exploration.

References

Carey, S. (1985). *Conceptual Change in Childhood*. Cambridge, MA: MIT Press/Bradford Books.

Barsalou, L. (1987) The instability of graded structure: Implications for the nature of concepts. In U.Neisser (Ed.), *Concepts and Conceptual Development:Ecological and Intellectual Factors inCategorization.* 101-140. New York:Cambridge University Press.

Barsalou, L. (1999). Perceptual Symbol Systems. *Behavioral and Brain Sciences,* 22: 577-660.

Carlson, G. (1977). Reference to kinds in English. Unpublished doctoral dissertation, University of Massachusetts, Amherst.

Devitt, M. & Sterelny, K (1999). *Language and Reality: An Introduction to the Philosophy of Language.* Second ed. MIT Press.

Fodor, J.A. (1975). *The Language of Thought*. New York: Thomas Y. Crowell.

Fodor, J.A. (1981) The present status of the innateness controversy. In Fodor, J. *Representations: Philosophical Essays on the Foundations of Cognitive Science.*

Fodor, J.A. (1998). Concepts: Where Cognitive Science Went Wrong. New York: Oxford University Press.

Fodor, J.A. (2008). LOT2: *The Language of Thought Revisited*. Oxford University Press.

Gelman, S. (2003). The essential child: Origins of essentialism in everyday thought. New York: Oxford University Press.

Gelman, S. A., & Bloom, P. (2007). Developmental changes in the understanding of generics. *Cognition, 105*, 166-183.

Gopnik, A. and Tenenbaum, J. B. (2007).Bayesian networks, Bayesian learning, and cognitive development. *Developmental Science 10*(3), 281-287.

Griffiths, T. L., Kemp, C., and Tenenbaum, J. B. (2008). Bayesian models of cognition. In Ron Sun (ed.), Cambridge Handbook of Computational Cognitive Modeling. Cambridge University Press.

Hume, David. (1748). *An Enquiry into Human Understanding*.

Keil, F. C. (1989). Concepts, kinds, and cognitive development. Cambridge, MA: MIT Press.

Kemp, C., Perfors, A., and Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Developmental Science 10*(3), 307-321.

Locke, J. (1690) *An Essay Concerning Human Understanding*.

Oved, I. (2009). *Baptizing Meanings for Concepts*. Doctoral Dissertation. Department of Philosophy. Rutgers, The State University of New Jersey. New Brunswick.

Oved, I. (2014) Hypothesis formation and testing in the acquisition of representationally simple concepts. *Philosophical Studies.*

Oved I. & Fasel I. (2011). Philosophically inspired concept acquisition for artificial general intelligence. *Proceedings of the 4th Conference of Artificial General Intelligence*.

Perfors, A, Tenenbaum, J. B., & Regier, T. (in press). The learnability of abstract syntactic principles. *Cognition.*

Prasada, S. (2000). Acquiring generic knowledge. Trends in Cognitive Sciences, 4, 66-72.

Prasada, S., & Dillingham, E. (2006). Principled and statistical connections in common sense conception. *Cognition, 99 (1)*.

Prasada, S., & Dillinghamn, E. (2009). Representation of principled connections: A window onto the formal aspect ofcommon sense conception. *Cognitive Science*, 33.

Prasada, S., Khemlani, s., Leslie, S.J., & Glucksberg, S. (2013). Conceptual distinctions amongst generics. Cognition, 126, 405-422.

Prinz, J. (2002). Furnishing the Mind: Concepts and their Perceptual Basis. MIT Press.

Rosch, E. (1978) Principles of categorization. In: E. Rosch & B. Lloyd, eds., *Cognition and Categorization*. Hillsdale, N.J.: Erlbaum Associates. 27-48.

Smith, E.E. & Medin, D.L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Xu, F. and Tenenbaum, J. B. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science 10*(3), 288-297.