Title
Dual Character Concepts

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN
1069-7977

Authors
Knobe, Joshua
Prasada, Sandeep

Publication Date
2011

Peer reviewed

Dual Character Concepts

Joshua Knobe (joshua.knobe@yale.edu)
Program in Cognitive Science, P.O. Box 208306
New Haven, CT 06520-8306 USA

Sandeep Prasada (sprasada@hunter.cuny.edu)
Department of Psychology, 695 Park Ave
New York, NY 10065 USA

Abstract

Four experiments provided evidence for a class of 'dual character concepts.' Such concepts are characterized in terms of both (a) concrete features and (b) abstract values. Three experiments found that when an object fulfills only one of these two criteria, it is judged to be a category member in one sense but not in another. A fourth experiment showed that dual character concepts also support a distinct form of normative judgment.

Keywords: Concepts; natural kinds; normativity.

Imagine a physics professor who spends her days writing out equations but who clings dogmatically to a certain theoretical perspective against all empirical evidence. Does this person genuinely count as a *scientist*?

In a case like this, one might feel that both answers are in some sense correct. It might therefore seem right to say:

(1) There is a sense in which she is clearly a scientist, but ultimately, if you think about what it really means to be a scientist, you would have to say that she is not a scientist at all.

Now suppose we come upon a person who has never been trained in formal experimental methods but who approaches everything in life by systematically revising her beliefs in light of empirical evidence. In a case of this latter type, it might seem appropriate to make the converse sort of statement:

(2) There is a sense in which she is clearly not a scientist, but ultimately, if you think about what it really means to be a scientist, you would have to say that she truly is a scientist.

To the extent that people do in fact show these patterns of intuition, we might conclude that they actually have two different characterizations of what it means to be a scientist – one in terms of concrete activities (conducting experiments, formulating theories, etc.), the other in terms of more abstract values (an impartial quest for empirical truth).

The implicit assumption in most work on conceptual representation seems to have been that concepts characterize members of a category in a single way – whether via the representation of a definition, a prototype, salient exemplars, or a theory (for a review, see Murphy, 2002). Could current approaches to conceptual representation accommodate concepts that provide two ways of characterizing their members? Or must they be modified to handle such concepts? Before these broader issues can be addressed, we need a better understanding of the concepts in question. The experiments in this paper are aimed at providing this information.

Dual character concepts

The experiments seek to demonstrate that there is a whole class of concepts which are represented via both (a) a set of concrete features and (b) a set of abstract values that the concrete features are seen as realizing. These two representations are intrinsically related, but they are nonetheless distinct, and they can sometimes yield opposing verdicts about whether a particular object counts as a category member or not.

We will argue that this pattern of intuitions can be found across a broad array of different concepts: SCIENTIST, ART, CRIMINAL, TEACHER, ROCK MUSIC, MOTHER, LOVE, and many others. These concepts, we suggest, differ fundamentally from the types of concepts that have been studied in the existing literature (e.g., from natural kind concepts). We will refer to them as *dual character concepts*.

Not all concepts, however, are dual character concepts. Take the concept BUS DRIVER. It would be odd to say something like (3) of a person who does not have any of the features normally associated with bus drivers.

(3) There is a sense in which she is clearly not a bus driver, but ultimately, if you think about what a bus driver really is, you would have to say that she truly is a bus driver.

This latter concept does not appear to provide an abstract way of characterizing a category. Similarly for a wide range of other concepts: BUS DRIVER, PHARMACIST, ACQUAINTANCE, RUSTLING NOISE, SECOND COUSIN, and so on. These concepts, we suggest, do not have a dual character. We will use them in our studies as control concepts.

The role of normative considerations

Dual character concepts provide two distinct ways of characterizing category members: one based on concrete features, the other based on what we have been calling 'abstract values.' The latter way of characterizing category

members implicates a normative dimension in the representation of dual character concepts. What is the nature of these normative representations?

It seems that they are importantly different from the normative representations that have been investigated within the literature on concepts. A number of studies have shown that normative judgments can impact judgments of typicality (Barsalou, 1985; Lynch, Coley & Medin, 2000). When people are thinking about typical instances of goal defined categories such as diet food, they tend to look for a food that is in some way ideal for dieting (Barsalou, 1985). Furthermore, experts' judgments of the typicality of natural kind categories such as trees are influenced by judgments about how desirable or ideal a tree is (Lynch, Coley & Medin, 2000; Burnett, Medin, Ross & Blok, 2005).

Dual character concepts require an additional normative dimension. This point comes out especially clearly when we consider cases in which two dimensions of normativity are found in a single concept. The concept *scientist*, for example, is associated with various concrete activities, and we can imagine a person who shows excellence in all of them (a talent for theory, experimental design, statistical analysis, etc.). We might praise such a person by saying:

(4) She is a good scientist.

This normative evaluation might play a certain role in intuitions about typicality, as predicted by existing theories. But there is also another, quite different dimension of normativity. Specifically, it might be thought that certain people embody, in their whole way of life, the broader values associated with the scientific enterprise. We could then praise a person who embodies these values by saying:

(5) She is a true scientist.

These two dimensions of normativity can sometimes come apart. For example, a person who does not have the relevant concrete skills but who nonetheless embodies throughout her life the abstract values that characterize science might not be a good scientist, but we could nonetheless praise her by saying 'She is a true scientist.'

Furthermore, these different dimensions of normativity appear to arise for different concepts. We can apply the notion of goodness across an enormous variety of concepts ('a good scientist,' 'a good coffee,' 'a good day'). By contrast, the second dimension of normativity seems to arise only for concepts in a more restricted class. A person might embody the values that characterize science and therefore be regarded as a 'true scientist,' or a painting might embody the values that characterize art and therefore be regarded as a 'true work of art,' but there are other cases in which this mode of thinking seems not to get a grip. A person might be highly skilled at driving buses and therefore be known as a good bus driver, but it seems hard to imagine how we could take a person to embody the broader values that characterize bus driving and therefore say of her: 'She is a true bus driver.'

In short, it appears our conceptual systems support at least two types of normative judgments and that dual character concepts support both types of normative judgments.

Hypotheses and overview of experiments

Our claim is that there is a distinctive class of concepts – the dual character concepts – which show two important characteristics:

> *Hypothesis 1*. Dual character concepts represent a system of abstract values which characterize members of the category. Thus, given any dual character concept *c*, it should be possible to make judgments about whether a given object realizes those values and can therefore be described by a sentence of the form 'That is a true *c*.'

> *Hypothesis 2*. Dual character concepts represent two distinct ways of characterizing members of a category: one in terms of relatively concrete features, the other in terms of the abstract values that these features realize. These two ways of characterizing members of the category allow people to make two independent assessments of category membership.

To conduct these experiments, we need a set of concepts hypothesized to have a dual character and a contrasting set of concepts that can be used as controls. For dual character concepts, we used SCIENTIST, ROCK MUSIC, LOVE, ARTIST, FRIEND, CRIMINAL, TEACHER, MOTHER, SOLDIER and POEM. For control concepts, we used BARTENDER, RUSTLING NOISE, IRRITATION, OPTICIAN, ACQUAINTANCE, CASHIER, PHARMACIST, SECOND COUSIN, CARJACKER and TABLE OF CONTENTS.

As a stimulus check, we conducted a brief study to verify that values play a role in characterizing the concepts we picked out as dual character concepts but not the control concepts. Twenty-five participants recruited through Amazon's Mechanical Turk were asked about all twenty of the concepts in random order. For each of the concepts, participants were told to imagine someone saying: 'He is a scientist [bartender, optician, etc.].' They were then told to imagine another person responding:

> I completely disagree. That person is not really an artist [bartender, optician, etc.] at all. In fact, if you think that he is really an artist [bartender, optician, etc.], I would have to say that you have some fundamentally wrong values.

The question for each item was whether this reference to values made sense. Participants marked their answers on a scale from 1 ('doesn't make sense') to 7 ('makes sense'). As predicted, the reference to values was judged to make more sense for the dual character concepts ($M = 3.8$) than for the control concepts ($M = 2.2$), ($t_1(24) = 6.87$, $p < .001$; $t_2(18) = 5.50$, $p < .001$). In fact, the distribution of means for the items was non-overlapping with the exception of one item. This suggests that the stimulus set correctly distinguished concepts that are characterized via abstract values and those

that do not. We used this set of items for generating the stimuli for all 4 experiments.

Experiment 1

Experiment 1 tests the hypothesis that dual character concepts support judgments not only about whether something is a 'good' category member but also whether it is a 'true' category member whereas control concepts only support the former type of normative judgment.

Method

For each of the dual character and control concepts, we generated a statement of the form *That is a good x* and *That is a true x*. The statements were presented with a 7-point scale whose ends were labeled *sounds weird* and *sounds natural*.

Twenty-nine English speakers participated via the Internet using Amazon's Mechanical Turk (AMT). Each participant received all 40 statements in a random order and rated how natural they sounded.

Results

The mean ratings are shown in Figure 1. Results were analyzed using 2x2 ANOVAs with concept type (dual character/control) and statement type (good/true) as factors. We report both participant (F_1) and item analyses (F_2).

The key prediction of an interaction between category type and statement type was confirmed ($F_1(1, 28)= 68.51$, $p <.001$; F_2 (1. 18) = 4.70. $p <.05$). As predicted, the interaction was due to significantly higher ratings for good statements as compared to true statements for the control categories (F_1 (1. 28) = 60.12. $p <001$; F_2 (1, 9) = 8.85, $p <.02$), but no difference between the two statement types for the dual character categories ($F_1 <1$; $F_2<1$).

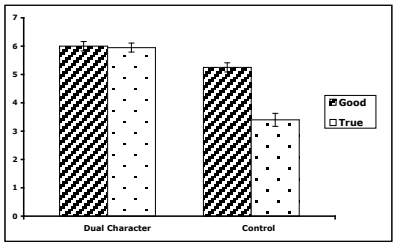

Figure 1. Mean ratings by condition for Expt. 1. (Error bars show SE mean.)

Discussion

As predicted, participants were willing to apply the word 'good' to both dual character concepts and control concepts ('good scientist,' 'good cashier'), but their use of the word 'true' was more restricted. They were willing to apply this word to dual character concepts ('true scientist') but not to control concepts ('true cashier'). This result provides some initial evidence for the hypothesis that dual character concepts differ in important respects from other concepts and, in particular, that they support a distinctive abstract form of normative judgment.

Experiment 2

Experiment 2 investigated the hypothesis that since dual character concepts provide two distinct ways of characterizing members of a category, they would allow two independent assessments of category membership. In this experiment, participants received a series of vignettes in which a category member (x) was described as possessing the concrete properties characteristic of a category (C), but lacking certain abstract normative properties. For example, the vignette for the dual character concept ARTIST described a person who creates paintings for a living but who has no real interest in creating work of deep aesthetic value and is simply trying to make money. Similarly, the vignette for the control concept PHARMACIST described a person who fills prescriptions for a living but who has no real interest in helping people and is simply trying to make money. After reading the vignettes, participants were asked to judge the truth of two statements. The first was what we will call the *member statement*: 'There is a sense in which this person is clearly a scientist [bartender].' The second was the *non-member statement*: 'Ultimately when you think about what it really means to be a scientist [bartender], you would have to say that this person is not truly a scientist [bartender].'

For dual character concepts, it was predicted that participants would agree with both statements. By contrast, for control concepts, it was predicted that participants would not see the concept as having any separate abstract sense and hence that they would agree only with the first statement.

Method

Thirty native English speakers participated in the experiment via the Internet using AMT. Vignettes were constructed for each concept used in Experiment 1 in the manner described above.

Each participant received all 20 vignettes in a random order and judged the truth of the 'member' and 'non-member' statements for each vignette on a 7 point scale.

Results

The mean ratings are shown in Figure 2. 2x2 ANOVAs with category type (dual character, control) and membership statement type (member, non-member) were performed.

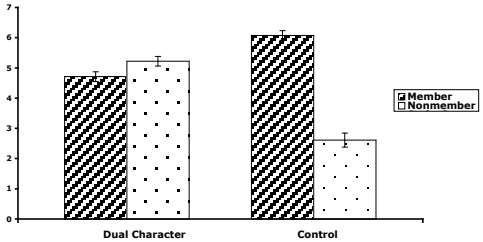

Figure 2. Mean ratings by condition for Expt. 2.

(Error bars show SE mean.)

The key prediction of an interaction between category type and statement type was confirmed ($F_1(1, 29)= 136.24$, $p <.001$; $F_2 (1. 18) = 59.60$. $p <.001$). As predicted, the interaction was due to significantly higher ratings for member statements as compared to non-member statements for the control categories ($F_1 (1. 29) = 138.63$. $p <001$; $F_2 (1, 9) = 71.63$, $p <.001$), but no difference between the two statement types for the dual character categories ($F_1 (1,29)= 2.61$, $p>.12$; $F_2(1,9)= 2.67$, $p>.14$).

Discussion

The results of Experiment 2 provide clear evidence that dual character concepts provide two bases for categorizing items as belonging in the category. Participants were willing to grant membership on the basis of concrete properties, but were also in agreement that those items that lacked the abstract normative characteristics of the category were not ultimately true members of the category. This contrasted with the control concepts where participants judged the object to be clearly a category member and did not see any more abstract sense in which one could say that, ultimately, it was not truly a category member at all. Thus dual character concepts allow participants to judge category membership either on the basis of the concrete properties or the abstract normative values that categorize the category. With the control categories, what you see is what you get.

Experiment 3

The fact that something other than readily observable concrete properties are relevant to category membership for dual character concepts is reminiscent of research on the representation of natural kind concepts (e.g. Keil, 1989; Gelman & Wellman, 1991). For natural kind concepts, the "something other" is understood to be hidden causes or essences. Could the judgments observed in the previous experiment be the same phenomena that have been observed for natural kind concepts? We do not think so. One important difference is that whereas dual character concepts provide two legitimate ways of characterizing and categorizing category members, natural kind concepts are usually understood to have only one legitimate way of determining category membership, with the easily observable superficial properties providing at best a quick and dirty heuristic for determining category membership.

Experiment 3 investigates whether dual character concepts and natural kind concepts differ in this manner by running an experiment like Experiment 2 with the addition of vignettes of natural kind categories. The natural kind vignettes were adapted from Keil (1989) and described concrete superficial characteristics of a given category, but lacking crucial underlying causal factors of that category.

A complex pattern of judgments across the three types of concepts was predicted. For control concepts, participants should focus on the concrete observable properties and ignore the more abstract values. Conversely, for the natural kind properties, they should focus on the hidden essence and ignore the concrete observable properties. For dual character concepts, participants should say that it can count as a category member in one sense while simultaneously not counting as a category member in another.

Method

Thirty native English speakers participated in the experiment via the Internet using AMT. In addition to the vignettes from Experiment 2, we included 10 vignettes of natural kind categories that were adapted from Keil (1989). These vignettes described things with the concrete superficial characteristics of a given category, but lacking the crucial underlying causal factors of that category. The crucial underlying causal factors described were those of a different category.

Results

The mean ratings given in each condition are shown in Figure 3. The key prediction of an interaction between category type and statement type for dual character and natural kind concepts was confirmed ($F_1(1, 29)= 74.23$, $p <.001$; $F_2 (1, 18) = 35.39$. $p <.001$). As predicted, the interaction was due to significantly higher ratings for non-member statements as compared to member statements for the natural kind categories ($F_1 (1, 29) = 127.79$. $p <001$; $F_2 (1, 9) = 114.29$, $p <.001$), but no difference between the two statement types for the dual character categories ($F_1 (1, 29)= 1.28$, $p >.27$; $F_2 < 1$).

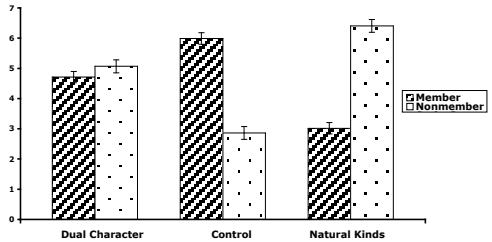

Figure 3. Mean ratings by condition for Expt. 3.
(Error bars show SE mean.)

The interaction between category type and statement type for natural kind and control concepts was also significant ($F_1(1, 29)= 206.22$, $p <.001$; $F_2 (1, 18) = 153.78$. $p <.001$). As predicted, the interaction was due to significantly higher ratings for non-member statements as compared to member statements for the natural kind categories ($F_1 (1, 29) = 127.79$. $p <001$; $F_2 (1, 9) = 114.29$, $p <.001$), but the opposite for the control concepts ($F_1 (1, 29)= 117.25$, $p<.001$; $F_2 (1, 9) = 55.70$, $p <.001$).

Discussion

Experiment 3 provides clear evidence that dual character concepts are represented differently than natural kind concepts. There is, however, a methodological issue one may raise concerning Experiments 2 & 3. Perhaps the differences between concept types was due to the vignettes

not being properly matched? We addressed this methodological concern in Experiment 4.

Experiment 4

In Experiment 4, we did away with the vignettes and simply asked participants to judge the extent to which statements of the following sort sounded weird/sounded ok to them:

> (6) There's a sense in which she is clearly a scientist/bartender, but ultimately, if you think about what it really means to be a scientist/bartender, you'd have to say that there is a sense in which she is not a scientist/bartender at all.

We predicted that such statements would sound fine for dual character concepts, but not for the control single character concepts.

If dual character concepts provide two ways of characterizing and thus categorizing items, it should also be possible to deny membership on the basis of concrete characteristics, but allow membership on the basis of the item embodying the abstract normative characteristics that characterize the category. Thus, we predict that participants should find statements of the following sort to sound fine for dual character concepts, but not the control concepts.

> (7) 'There's a sense in which she is clearly not a scientist/bartender, but ultimately, if you think about what it really means to be a scientist/bartender, you'd have to say that there is a sense in which she is a true scientist/bartender after all.

Method

The concepts used in Experiments 1 & 2 were used to construct statements of the form of (X) and (Y). We refer to these as *concrete-only* and *abstract-only statements*.

Thirty native English speakers participated in the experiment via the Internet using AMT. Participants were instructed to rate the extent to which the sentences sounded ok/sounded bad. Each participant received all 40 items in a different random order.

Results

The mean ratings by condition are shown in Figure 4. As predicted, there was a main effect of concept type with higher ratings for dual character concepts ($F_1(1, 29)= 91.34$, $p <.001$; $F_2 (1, 18)=105.83$, $p <.001$). There was no effect of statement (F_1& $F_2 < 1$), or interaction (F_1& $F_2 < 1$).

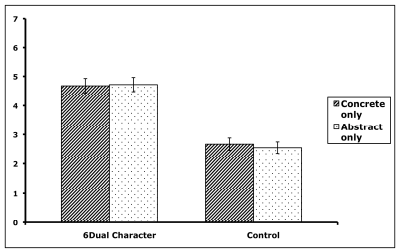

Figure 4. Mean ratings by condition for Expt. 4. (Error bars show SE mean.)

Discussion

The results of this experiment suggest that the effects seen in Experiments 2 and 3 were not artifacts due to potentially unmatched vignettes. On the contrary, even when one omits the vignettes, participants still think that certain sentences sound right with dual character concepts but not with control concepts.

Furthermore the fact that both types of statements were rated highly for dual characteristic concepts shows that the concrete and abstract criteria can come apart in either direction. Just as it is possible to fulfill the concrete criteria without fulfilling the abstract ones, it is possible to fulfill the abstract criteria without fulfilling the concrete ones.

General Discussion

Four experiments found that dual character concepts differed from the control concepts across a number of different measures.

First, participants were willing to apply the adjective 'true' with dual character concepts ('true scientist') but not with control concepts ('true cashier').

Perhaps more tellingly, when participants were given vignettes about objects that had all the concrete features that would be stereotypically associated with a category but nonetheless lacked the abstract values that characterize the category, they were willing to use sentences of the form:

> (8) There is a sense in which she is clearly a scientist, but ultimately, if you think about what it really means to be a scientist, you would have to say that she is not a scientist at all.

However, participants were not willing to use sentences of this form with either control concepts or natural kind concepts. Indeed, even in the absence of any vignette, they indicated that sentences like this make sense with dual character concepts but not with control concepts.

Finally, in cases of dual character concepts, people were able to make sense of the converse sentence:

> (9) There is a sense in which she is clearly not a scientist, but ultimately, if you think about what it really means to be a scientist, you would have to say that she truly is a scientist.

However, participants indicated that such sentences could not be applied in cases of control concepts.

All in all, then, the experimental evidence indicates that dual character concepts do indeed differ from other concepts in systematic ways. We will be focusing here on two claims about the nature of this difference: (a) that each dual character concept provides two distinct criteria for category membership; (b) that one of these criteria involves certain kinds of abstract values.

Distinct criteria

One striking aspect of dual character concepts is that people are willing to say that a single object can fall under such a concept in one sense while not falling under the concept in another. How is this duality to be understood?

One way to get a handle on this phenomenon is to compare it with cases in which people have only a single system of criteria but these criteria allow different degrees of category membership. Take the concept HOT CHOCOLATE. It might happen that some objects fulfill the criteria for this concept to a high degree, others do not fulfill them at all, and some stand right at the border, such that a person could say: 'Loosely speaking, this is hot chocolate, but strictly speaking, it is not hot chocolate.' Yet one should not therefore conclude that people have two distinct systems of criteria for hot chocolate. Rather, it seems they have a single system of criteria, and although certain objects fulfill those criteria perfectly, others only fulfill them to a certain degree.

The results of the present studies, however, suggest that something more is going on in the case of dual character concepts. To say that an individual is a 'true scientist' is not simply to say that this individual exhibits the characteristic features of scientists to an unusually high degree. After all, as we found in Experiment 4, people are willing to say that there might be a certain sense in which an individual counts as a true scientist even if there is a sense in which this individual does not count as a scientist at all.

So what we see arising in these cases appears to be something more fundamental. It seems that people actually are adopting two distinct systems of criteria, such that a single object can fulfill either one of the criteria while failing to fulfill the other.

Abstract values

A second striking aspect of dual character concepts is the role played by abstract values. People appear to understand the concept SCIENTIST in terms of an impartial quest for empirical knowledge, ROCK MUSIC in terms of raw emotional intensity, and so on. When an object fails to embody these abstract values, participants conclude that even if it has the concrete features associated with a given category, there is still a sense in which it does not count as a genuine category member.

The phenomenon at work here seems to show a certain structural similarity to the one we find in natural kind concepts. Faced with a natural kind concept like SKUNK, people might observe numerous different superficial features (stripedness, smelliness, etc.), but they also conclude that these superficial features are united by all having the same underlying causes. These underlying causes are then seen as the true criteria for category membership. Likewise, in the case of a dual character concept like scientist, people might observe numerous different concrete activities (conducting experiments, analyzing data, etc.), but they conclude that all of these concrete activities are ways of realizing the same abstract values. In an analogous fashion, these abstract values are treated as criteria for category membership.

Yet, despite these structural similarities, there are also important respects in which dual character concepts differ from natural kind concepts. Natural kinds are characterized in terms of underlying causes (e.g., DNA), whereas dual character concepts are characterized in terms of abstract values (e.g., raw emotional intensity). The hidden essence of a natural kind concept is often entirely unknown, whereas the abstract values of a dual character concept are typically known. And, as Experiment 3 shows, participants judge that membership in a natural kind is independent of concrete superficial features, whereas participants think that there is a sense in which concrete features can be sufficient for membership in a dual character category.

To capture the distinctive properties of dual character concepts, it may be necessary to extend the theories developed for natural kind concepts in a number of directions. Above all, it may be necessary to generalize these theories so that they apply not only to causal relations but also to normative relations. Such a generalization may at first seem surprising, but a surge of recent research has indicated that normative considerations can impact people's cognition in a wide array of domains, including everything from causal judgment to theory-of-mind (for a review, see Knobe 2010). In combination with other recent findings (Prasada & Dillingham 2006; 2009), the present results suggest that this pervasive impact of normativity extends to conceptual representation as well.

References

Barsalou, L.W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 629-654.

Gelman, S.A. & Wellman, H.M. (1991). Insides and essences: Early understanding of the non-obvious. *Cognition, 38,* 213-244.

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

Knobe, J. (2010). Person as Scientist, Person as Moralist. *Behavioral and Brain Sciences*. 33: 315-329.

Lynch, E. B., Coley, J. D., & Medin, D. L. (2000). Tall is typical: Central tendency, ideal dimensions, and graded category structure among tree experts and novices. *Memory & Cognition, 28,* 41- 50.

Murphy, G.L. (2002). *The big book of concepts.* MIT Press, Cambridge, MA.

Prasada, S., & Dillingham, E. (2006). Principled and statistical connections in common sense conception. *Cognition, 99*, 73-112.

Prasada, S., & Dillingham, E. (2009). Representation of principled connections: A window onto the formal aspect of common sense conception. *Cognitive Science*, *33*, 401-448.